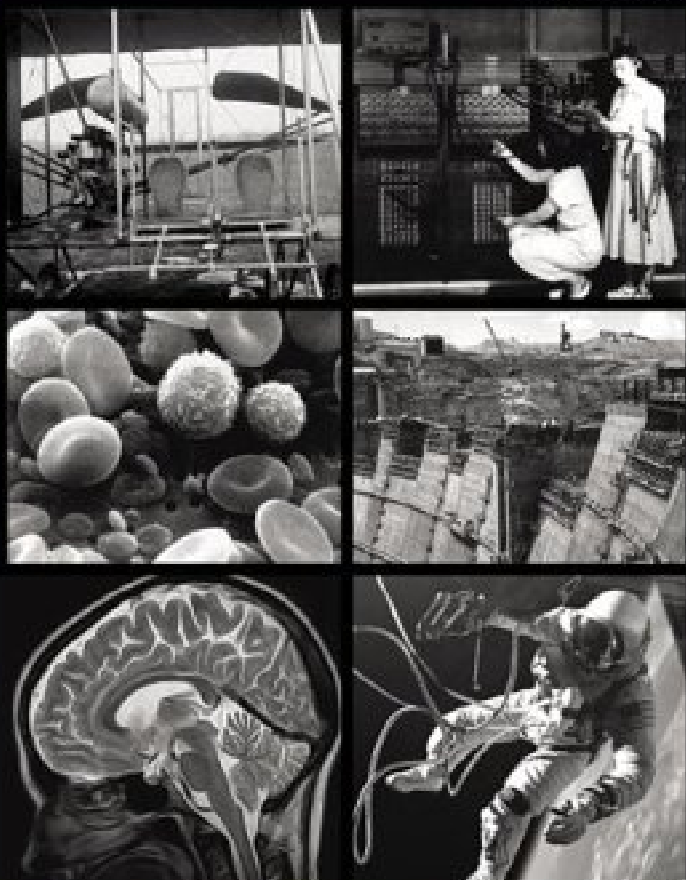
ENCYCLOPEDIA OF 20TH-CENTURY

TECHNOLOGY

Volume 1, A-L



Colin A. Hempstead, Editor ▪ William E. Worthington, Jr. Associate Editor

ENCYCLOPEDIA OF 20TH-CENTURY
TECHNOLOGY

Volume 1
A–L

Colin A. Hempstead, Editor
William E. Worthington, Jr., Associate Editor

ROUTLEDGE
NEW YORK AND LONDON

Also available as a printed book
see title verso for ISBN details

ENCYCLOPEDIA OF 20TH-CENTURY
TECHNOLOGY

ENCYCLOPEDIA OF 20TH-CENTURY
TECHNOLOGY

Volume 1
A–L

Colin A. Hempstead, Editor
William E. Worthington, Jr., Associate Editor

ROUTLEDGE
NEW YORK AND LONDON

Published in 2005 by
Routledge
270 Madison Avenue
New York, NY 10016
www.routledge-ny.com

Published in Great Britain by
Routledge
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN, UK
www.routledge.co.uk

Copyright © 2005 by Routledge

Routledge is an imprint of the Taylor & Francis Group.

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's
collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage and retrieval system, without permission in writing from the publisher.

Library of Congress Cataloging-in-Publication Data

Encyclopedia of 20th-century technology / Colin A. Hempstead, editor; William E. Worthington, associate editor.
p. cm.
Includes bibliographical references and index.
ISBN 1-57958-386-5 (set : alk. paper)—ISBN 1-57958-463-2 (vol. 1 : alk. paper)—
ISBN 1-57958-464-0 (vol. 2 alk. paper)
1. Technology—Encyclopedias. I. Hempstead, Colin. II. Worthington, William E., 1948–
T9.E462 2005
603—dc22

ISBN 0-203-99699-2 Master e-book ISBN

Advisers

Dr. Jon Agar, Department of History and Philosophy of Science, University of Cambridge

Professor Janet Bainbridge, Chief Executive, EPICC (European Process Industries Competitiveness Centre), Teesside

Dr. Hans Joachim Braun, Universität der Bundeswehr Hamburg

Dr. Robert Bud, Principal Curator of Science, Science Museum, London

Dr. Michael Duffy, formerly Department of Engineering, University of Sunderland

Dr. Slava Gerovitch, Dibner Institute for the History of Science and Technology, MIT

Dr. Ernst Homburg, Department of History, University of Maastricht

Dr. Sally Horrocks, Department of Economic and Social History, University of Leicester

R. Douglas Hurt, Professor and Director, Graduate Program in Agricultural History and Rural Studies, Department of History, Iowa State University

Dr. Peter Morris, Science Museum, London

Professor John Pickstone, Centre for the History of Science, Technology and Medicine, University of Manchester

Keith Thrower, former Technical Director at Racal Electronics, UK

Contents

List of Entries

Thematic List of Entries

Television, Radio, Audio Recording

Warfare

Editor's Preface

All editors of encyclopedias are faced with the problem of what to include. Even if the title is agreed and the numbers of volumes and pages have been decided, the sum of possible entries could be very large. In the case of the *Encyclopedia of 20th-Century Technology*, the editor decided that in order to construct a logical and consistent set of entries it was necessary to adopt what could be described as an analytic framework. During the 20th century a plethora of manufactured articles have appeared for which the real costs have continuously fallen. The products in industrialized societies have become universal, and many of the good ones are within the reach of a large proportion of humanity. In keeping with this democratic trend of the century it was decided that people and their experiences with technology should be central to the encyclopedia. Readers are urged to read the entries in the light of the humanistic core.

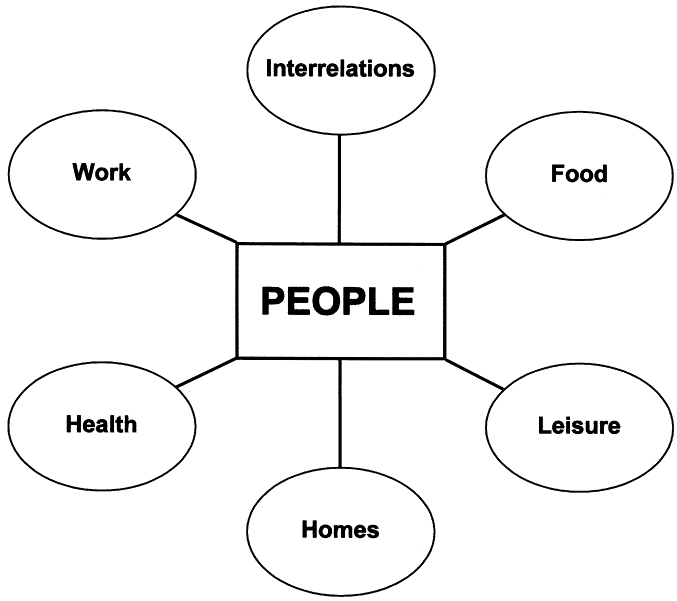
An examination of people and their lives led to six broad, related areas of society from which the four hundred entries that comprise these volumes could be derived. The type of analysis carried out is indicated in the diagrams on the next page. The first shows the six basic areas; the second diagram is an outline of the detailed application for the category FOOD. Five or six levels of analysis allowed the definition of headers that provided the individual entries. Of course, entries could be found in two or more basic areas or could be related to others: entries in refrigerating in the domestic situation as found in food preservation would lead to entries in the technology of refrigeration *per se*. Thus the contents were defined.

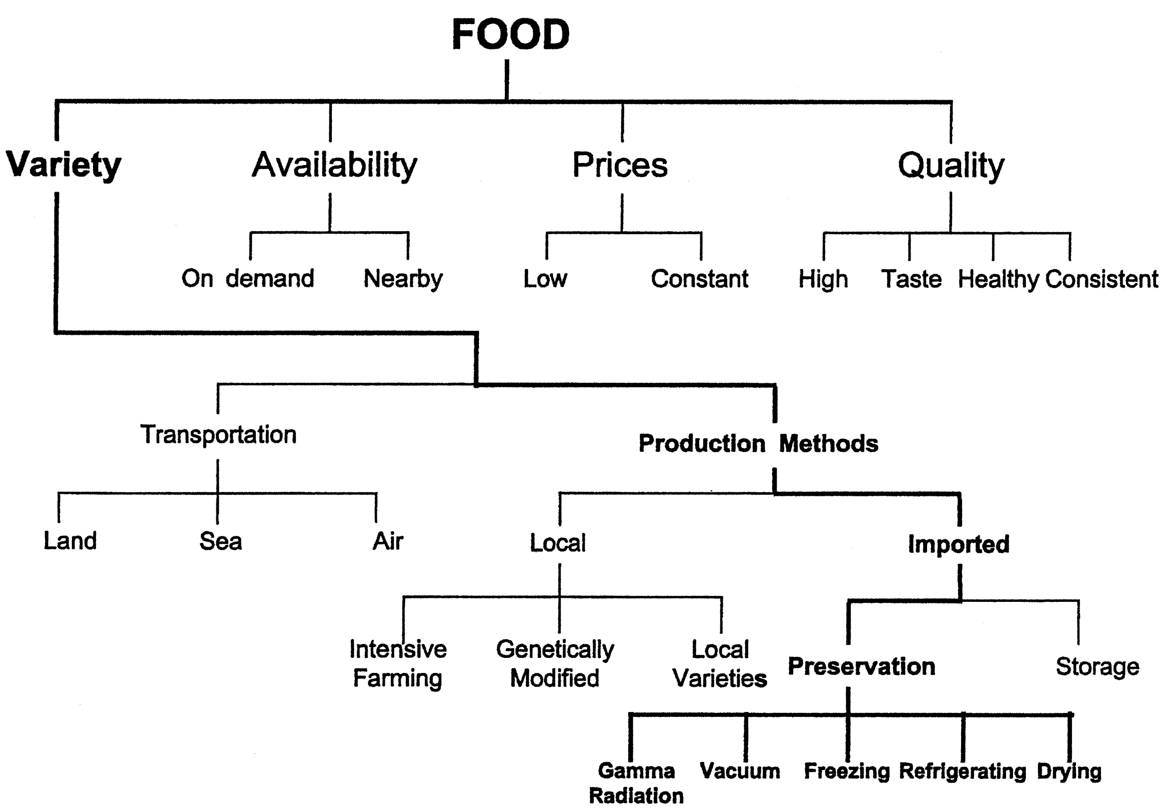
The encyclopedia contains two types of entries. The greatest number of entries are of 1000 words, and as far as possible these standard entries are devoid of interpretation. Nevertheless, it is recognized that all history is redolent of the era in which it is constructed, and this encyclopedia is of its own particular society, that of Western industrial. The factual nature of the standard entries is leavened by longer essays in which historical and interpretative themes are explored. Among other things, these essays describe and analyze the relationship between society and technology, touch on the modern debates on the nature of the history of technology of history, and relate what people expect of the products of modern industrial civilisation.

The encyclopedia is concerned with 20th-century technology but not with 20th-century inventions. The technologies included are those that had an impact on the mass of the population in industrial societies. So many technologies invented in the 19th century did not begin to impinge markedly on many lives until the middle of the 20th century, so they are considered to be of the 20th century. Similarly, many products in the constructed world are old conceptions, transformed by modern materials or production methods. They have found a place in the encyclopedia. The inclusion of pre-20th-century products compares with the exclusion of recently developed technologies that have yet to have any effect on the mass of the public. However, the encyclopedia is not intended to be futuristic. In the 20th century, scientific engineering came to majority, and many if not all the products of modern technology can be seen to be the results of science. However, there are no entries that discuss science itself. Within the essays, however, science as science related to each subject is described.

Even with four hundred entries, the encyclopedia is not canonical, and gaps will be noted. However, the standard entries, the interpretative essays, and the lists of references and further reading suggestions allow readers to appreciate the breadth and depth of the technology of the 20th century.

Colin Hempstead


PEOPLE
Interrelations
Work
Food
Health
Leisure
Homes


FOOD
Variety
Availability
Prices
Quality
On demand
Nearby
Low
Constant
High
Taste
Healthy
Consistent
Transportation
Production Methods
Land
Sea
Air
Local
Imported
Intensive Farming
Genetically Modified
Local Varieties
Preservation
Storage
Gamma Radiation
Vacuum
Freezing
Refrigerating
Drying

Associate Editor's Preface

Technology is a vital subject. It grows continuously. New technologies are introduced, existing technologies evolve, and the outmoded are abandoned. Looking dispassionately at technology, it is always exciting, for it is the product of human ingenuity. For the purposes of this encyclopedia, we felt it could not and should not be discussed devoid of its human element. It is breathtaking to consider the panoply of developments which occurred during the last century, but it is necessary to recall that these developments did not take place in isolation. It was our desire to see that events, where possible, were described in context. Thus, you will find names, places, dates, and events critical to the development of a particular technology. The reader will note that some entries contain a surprising amount of information on 19th-century events. This was appropriate, for some 20th-century technologies were firmly rooted in that earlier time and can be best understood in light of the past. To avoid a deadly dull recitation of formulae and regurgitation of dry facts, we sought to give the reader the broadest possible picture.

The encyclopedia was created for the lay reader and students as well as for historians of science and technology. In light of this, we attempted to minimize the use of the jargon that tends to grow around some technologies. Although many of the subjects are highly technical, our belief was that even complicated subjects could be rendered in such a way as to make them comprehensible to a wide audience. In the same way that an electrical engineer might need explanations when encountering genetic terminology, students and non-specialists will also appreciate the clarification. Because of the pervasiveness of the subjects in all facets of our lives, the encyclopedia should be a handy reference tool for a broad range of readers. Our aim was to make the subjects, which many of us deal with daily and do not necessarily grasp completely, readily understood with a minimum need for additional reference. However, should the reader wish to delve further into any particular subject, our expert authors have provided a selection of further bibliographic readings with which to begin.

The scope of the encyclopedia is intended to be international. Discussions were to be as inclusive as possible and avoid focus solely on the events of any one country. Nonetheless, some skewing was unavoidable due simply to the prodigious number of developments that have taken place in some countries.

William E. Worthington, Jr.

Acknowledgments

A host of workers and authors contributed to this encyclopedia, and I wish to extend my thanks to every person without whom these volumes would be stillborn. My particular thanks are offered to Gillian Lindsey of Routledge. Gillian conceived the idea of an encyclopedia of 20th-century technology, and appointed me the editor of the work in 2000. Her energy and ideas were legion, although she glossed over the amount of work for me! However, the editorship was rewarding, offering the possibility of producing a worthwhile publication with academic colleagues from around the globe. The selection of technologies and of particular subjects suggested by Gillian and me were critiqued and extended by our advisers. Their contributions, drawn from their specialist knowledge and scholarship, were invaluable. When circumstances forced my withdrawal from the active editorship, William Worthington, then with the National Museum of American History in Washington, stepped into the hot seat. To William I give my heartfelt thanks.

Finally I acknowledge the publishers and the 20th century which presented all of us with the opportunity to examine and extol some of the content and effects of modern technology. Nevertheless, the encyclopedia is partial, and any omissions and shortcomings are mine.

Colin Hempstead

My thanks go to Gillian Lindsey for presenting me with the challenge of filling the void left by Colin's departure. However, the prospect of assuming a role in a project already well under way and natural differences in approach and style were concerns. Nonetheless, the final third of the encyclopedia was crafted in such a way that it blends seamlessly with the sections completed under Colin's careful guidance. This was due in no small part to the untiring efforts of Sally Barhydt, and to her I extend sincere thanks.

William E. Worthington, Jr.

Contributors

W.A. Addis, Buro Happold, Middlesex, United Kingdom.
Aaron Alcorn, Cleveland Heights, Ohio, USA.
K.W. Allen, Joining Technology Research Center, Oxford Brookes University, United Kingdom.
Paul Anastas, White House Office of Science and Technology, National Security and International Activities Division, USA.
Joe Anderson, Agricultural History and Rural Studies Program, Iowa State University, USA.
Stuart Anderson, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, United Kingdom.
Rachel A. Ankeny, Unit for History and Philosophy of Science, University of Sydney, Australia.
Peter Asaro, Departments of Computer Science and Philosophy, University of Illinois at Champaign, USA.
Glen Asner, Department of History, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
Janet Bainbridge, School of Science and Technology, University of Teesside, and Chief Executive of EPICC (European Process Industries Competitiveness Centre), United Kingdom.
Guy D. Ball, Product Information, Unisys, Tustin, California, USA.
Susan B. Barnes, Department of Communication and Media Studies, Fordham University, New York, USA.
Keith Bartle, Department of Chemistry, University of Leeds, United Kingdom.
Donald R Baucom, Department of Defense, Ballistic Missile Defense Organization, USA.
Reinhold Bauer, Universität der Bundeswehr Hamburg, Germany.
Joyce E. Bedi, National Museum of American History, Lemelson Center for the Study of Invention and Innovation, USA.
Randal Beeman, Archives Director, Bakersfield College, California, USA.
Hal Berghel, Department of Computer Science, University of Nevada at Las Vegas, USA.
Beverly Biderman, Adaptive Technology Resource Centre, University of Toronto, Canada.
David I. Bleiwas, U.S. Geological Survey, Reston, Virginia, USA.
F.K. Boersma, Department of Culture, Organization and Management, Vrije Universiteit Amsterdam, Netherlands.
James Bohning, Center for Emeritus Scientists in Academic Research (CESAR), Lehigh University, Bethlehem, Pennsylvania, USA.
Brian Bowers, Engineering Historian and Writer, Retired Senior Curator, Science Museum, London, United Kingdom.
Hans-Joachim Braun, Universität der Bundeswehr Hamburg, Germany.
Catherine Brosnan, United Kingdom.
David J. Brown, Ove Arup & Partners, London, United Kingdom.
Louis Brown, Department of Terrestrial Magnetism, Carnegie Institution of Washington, USA.
Nik Brown, Science and Technology Studies Unit, University of York, United Kingdom.
Timothy S. Brown, Department of History, Pomona College, California, USA.
Robert Bud, Head of Research (Collections), Science Museum, London, United Kingdom.
William L. Budde, Office of Research and Development, U.S. Environmental Protection Agency.
Ian Burdon, Head of Sustainable Energy Developments, Energy and Utility Consulting, PB Power Ltd., Newcastle, United Kingdom.
Larry Burke, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
Russell W. Burns, Retired Professor, Nottingham, United Kingdom.
Michael Bussell, London, United Kingdom.
J. Stewart Cameron, St. Thomas' and Guy's Hospital, London, United Kingdom.

Rodney P. Carlisle, History Associates Incorporated, Rockville, Maryland, USA.
Stéphane Castonguay, Département des sciences humaines, Université du Québec á Trois-Rivières, Canada.
Carol A. Cheek, Rensselaer Polytechnic Institute, Troy, New York, USA.
Dennis W. Cheek, John Templeton Foundation, Radnor, Pennsylvania, USA.
Mark Clark, General Studies Department, Oregon Institute of Technology, USA.
Noel G. Coley, Department of the History of Science, Technology and Medicine, Open University, Milton Keynes, United Kingdom.
Paul Collins, Ironbridge Institute, University of Birmingham, United Kingdom.
Yonina Cooper, University of Nevada at Las Vegas, USA.
Peter Copeland, National Sound Archives, British Library, United Kingdom.
Anthony Coulls, Formerly of the Museum of Science and Industry in Manchester, United Kingdom.
Jennifer Cousineau, Charlottesville, Virginia, USA.
Trevor Cox, Acoustics Research Center, University of Salford, Greater Manchester, United Kingdom.
Jennifer Croissant, Program on Culture, Science, Technology, and Society, University of Arizona, USA.
Donard de Cogan, School of Information Systems, University of East Anglia, Norwich, United Kingdom.
Guillaume de Syon, Department of History, Albright College, USA.
Marc J. de Vries, Philosophy and Social Sciences, Eindhoven University of Technology, Netherlands.
Andrew Dequasie, Pownal, Vermont, USA.
Maggie Dennis, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Panos Diamantopoulos, School of Engineering, University of Sussex, United Kingdom.
John Dolan, Retired, Nobel Division ICI, United Kingdom.
Michael Duffy, Lancashire, United Kingdom; formerly of the Department of Engineering, University of Sunderland.
Charles Duvall, Bandwidth 9, Duluth, Georgia, USA.
Matthew Eisler, Department of History and Classics, University of Alberta, Edmonton, Canada.
Boelie Elzen, University of Twente, The Netherlands.
Linda Endersby, Thomas A. Edison Papers, Rutgers University, New Jersey, USA.
Ted Everson, Institute for the History and Philosophy of Science and Technology, University of Toronto, Canada.
Jim Farmer, Chief Technical Officer, Wave 7 Optics, Alpharetta, Georgia, USA.
David L. Ferro, Computer Science Department, Weber State University, Ogden, Utah, USA.
Mark Finlay, Department of History, Armstrong Atlantic State University, Savannah, Georgia, USA.
Gerard J. Fitzgerald, Department of History, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
Amy Foster, Ph.D. Candidate, Auburn University, Alabama, USA.
Philip L. Frana, Charles Babbage Institute, University of Minnesota, USA.
Philip J. Gibbon, Temple University, Philadelphia, Pennsylvania, USA.
Bruce Gillespie, Braamfontein, South Africa.
Julia Chenot GoodFox, University of Kansas, USA.
David Grier, Department of History, Center for History of Recent Science, George Washington University, Washington, D.C., USA.
Reese Griffin, 2 Stroke International Engine Co., Beaufort, South Carolina, USA.
Eric Grove, Center for Security Studies, University of Hull, United Kingdom.
David Haberstich, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
John Hamblin, Medical Doctor, Essex, United Kingdom.
Jennifer Harrison, Department of Information Technology, College of William & Mary, Williamsburg, Virginia, USA.
Ulf Hashagen, Deutsches Museum, Germany.
Peter Hawkes, CEMES-CNRS, Toulouse, France.
Stephen Healy, School of Science and Technology Studies, University of New South Wales, Sydney, Australia.
David Healy, North Wales Department of Psychological Medicine, University of Wales College of Medicine, Bangor, United Kingdom.
Colin Hempstead, Darlington, United Kingdom; formerly Reader in History of Science and Technology, University of Teesside.
Martin Hempstead, Corning Photonics Technology, Corning Inc., New York, USA.

Klaus Hentschel, Institut für Philosophie, University of Berne, Switzerland.
Arne Hessenbruch, Dibner Institute, Massachusetts Institute of Technology, USA.
Robert D. Hicks, Chemical Heritage Foundation, Philadelphia, Pennsylvania, USA.
Roger Howkins, Ove Arup & Partners, London, United Kingdom.
Peter J. Hugill, Department of Geography, Texas A&M University, USA.
Merritt Ierley, Sussex, New Jersey, USA.
Mary Ingram, Department of Sociology, University of California, Santa Barbara, USA.
Muzaffar Iqbal, Center for Islam and Science, Canada.
Ann Johnson, Department of History, Fordham University, New York, USA.
Sean Johnston, Science Studies, University of Glasgow, United Kingdom.
Suzanne W. Junod, History Office, U.S. Food and Drug Administration.
David Kaplan, Biomedical Engineering, Tufts University, Boston, Massachusetts, USA.
Christine Keiner, Science, Technology, and Society Department, Rochester Institute of Technology, New York, USA.
Karen D. Kelley, U.S. Geological Survey, Reston, Virginia, USA.
David Kirsch, Smith School of Business, University of Maryland, USA.
Timothy Kneeland, Department of History and Political Science, Nazareth College, Rochester, New York, USA.
Ramunas A. Kondratas, Division of Science, Medicine, and Society, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Helge Kragh, History of Science Department, University of Aarhus, Denmark.
John Krige, School of History, Technology, and Society, Georgia Institute of Technology, USA.
Alex Law, Department of Sociology, University of Abertay, Dundee, United Kingdom.
Michal Lebl, Illumina, Inc. & Spyder Instruments, Inc., San Diego, California, USA.
Tim LeCain, Department of History, Montana State University, USA.
Trudy Levine, Computer Science Department, Fairleigh Dickinson University, Teaneck, New Jersey, USA.
Gillian Lindsey, Writer and Editor, New York, USA.
Eric Livo, U.S. Geological Survey, Reston, Virginia, USA.
Henry Lowood, Curator, History of Science and Technology Collections, Stanford University, California, USA.
Juan Lucena, Liberal Arts and International Studies, Colorado School of Mines, USA.
Harro Maat, Technology & Agrarian Development (TAO), Wageningen University, Netherlands.
Alex Magoun, Executive Director, David Sarnoff Library, Princeton, New Jersey, USA.
A.M. Mannion, Department of Geography, University of Reading, United Kingdom.
J. Rosser Matthews, History Office, National Institutes of Health, Bethesda, Maryland, USA.
W. Patrick McCray, Center for History of Physics, American Institute of Physics, College Park, Maryland, USA.
Ian C. McKay, Department of Immunology and Bacteriology, University of Glasgow, United Kingdom.
Shelley McKellar, Department of History, University of Western Ontario, London, Canada.
Dennis McMullan, London, United Kingdom.
Kenneth Mernitz, History and Social Studies Department, Buffalo State College, New York, USA.
Lolly Merrell, Paonia, Colorado, USA.
Andre Millard, Department of History, University of Alabama at Birmingham, USA.
Carl Mitcham, Liberal Arts and International Studies, Colorado School of Mines, USA.
Susan Molyneux-Hodgson, Department of Sociological Studies, University of Sheffield, United Kingdom.
Gijs Mom, Foundation for the History of Technology, Technical University of Eindhoven, Netherlands.
John Morello, Department of General Education, DeVry Institute of Technology, Addison, Illinois, USA.
Peter Morris, Science Museum, London, United Kingdom.
Robin Morris, Retired Lecturer, West Malvern, United Kingdom.
David L. Morton, Tucker, Georgia, USA.
Susan Mossman, Senior Curator, Science Museum, London, United Kingdom.
Karel Mulder, Technology Assessment Group, Delft University of Technology, Netherlands.
Peter Myers, Department of Chemistry, University of Leeds, United Kingdom.
Francis Neary, Centre for the History of Science, Technology and Medicine, University of Manchester, United Kingdom.

Caryn E. Neumann, Department of History, The Ohio State University, USA.
William O'Neil, CNA Corporation, Alexandria, Virginia, USA.
Andrew Panay, Department of Sociology, University of Abertay, Dundee, United Kingdom.
Kayhan Parsi, Neiswanger Institute for Bioethics and Health Policy, Stritch School of Medicine, Loyola University of Chicago, USA.
Mike Pavelec, Department of History, The Ohio State University, USA.
Niocola Perrin, Nuffield Council on Bioethics, London, United Kingdom.
James Perry, Formerly Strategic Assessment Center, SAIC, USA.
John Pfotenhauer, Applied Superconductivity Center, University of Wisconsin-Madison, USA.
Robert Raikes, Meko Ltd., United Kingdom.
Thomas W. Redpath, Department of Biomedical Physics and Bioengineering, University of Aberdeen, United Kingdom.
Antoni Rogalski, Institute of Applied Physics, Military University of Technology, Warsaw, Poland.
David Rose, Retired Otorhinolaryngology Consultancy, Stockport NHS Trust, Manchester, United Kingdom.
Paul Rosen, Department of Sociology, University of York, United Kingdom.
Robin Roy, Department of Design and Innovation, Open University, United Kingdom.
Pedro Ruiz-Castell, St. Cross College, University of Oxford, Valencia, Spain.
Robert W. Rydell, Department of History, Montana State University, USA.
Nicholas Saunders, London, United Kingdom.
Roger Scantlebury, Integra SP, London, United Kingdom.
Jessica R. Schaap, Policy and Communications, NewMediaBC, Canada.
Elizabeth Schafer, Loachapoka, Alabama, USA.
Thomas Schlich, Institut fuer Geschichte der Medizin, Albert-Ludwigs-Universitaet, Freiburg, Germany.
Jeff Schramm, History Department, University of Missouri-Rolla, USA.
Stuart Shapiro, Senior Information Security Scientist, The MITRE Corporation, Bedford, Massachusetts, USA.
G. Terry Sharrer, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Duncan Shepherd, School of Engineering, University of Birmingham, United Kingdom.
John K. Smith, Department of History, Lehigh University, Bethlehem, Pennsylvania, USA.
John S. Sobolewski, Computer & Information Research and Technology, University of New Mexico, USA.
Lawrence Souder, Department of Culture and Communication, Drexel University, Philadelphia, Pennsylvania, USA.
James Steele, National Physical Laboratory, United Kingdom.
Carlene Stephens, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Christopher Sterling, George Washington University, Washington, D.C., USA.
Jack Stilgoe, University College London, United Kingdom.
Anthony Stranges, Department of History, Texas A & M University, USA.
James Streckfuss, College of Evening and Continuing Education, University of Cincinnati, USA.
Rick Sturdevant, HQ AFSPC/HO, Peterson, Colorado, USA.
Eric G. Swedin, Computer Science Department, Weber State University, Ogden, Utah, USA.
Derek Taylor, Altechnica, Milton Keynes, United Kingdom.
Ernie Teagarden, Professor Emeritus, Dakota State University, Madison, South Dakota, USA.
Jessica Teisch, Department of Geography, University of California, Berkeley, USA.
Thom Thomas, Halmstad University, Sweden.
Lana Thompson, Florida Atlantic University, Boca Raton, USA.
James E. Tomayko, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
Anthony S. Travis, Sidney M. Edelstein Center for History and Philosophy, Hebrew University, Jerusalem, Israel.
Simone Turchetti, Centre for History of Science, Technology and Medicine, University of Manchester, United Kingdom.
Steven Turner, National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Aristotle Tympas, Department of Philosophy and History of Science, University of Athens, Panepistimioupolis, Greece.

Eric v.d. Luft, Historical Collections, Health Sciences Library, State University of New York Upstate Medical University, Syracuse, USA.
Helen Valier, Centre for History of Science, Technology and Medicine, University of Manchester, United Kingdom.
Peter Van Olinda, Power Systems Engineer, New York, USA.
Colin Walsh, Medical Physics and Bioengineering Department, St. James's Hospital, Dublin, Ireland.
John Ward, Senior Research Fellow, Science Museum, London, United Kingdom.
Frank Watson, Reynolds Metals Company, Richmond, Virginia, USA.
David R. Wilburn, U.S. Geological Survey, Reston, Virginia, USA.
Mark Williamson, Kirkby Thore, The Glebe House, United Kingdom.
Duncan Wilson, Centre for History of Science, Technology and Medicine, University of Manchester, United Kingdom.
Frank Winter, Curator, Rocketry, National Air and Space Museum, Smithsonian Institution, Washington D.C., USA.
Bob Wintermute, Scholar in Residence, Army Heritage Center Foundation; Doctoral Candidate, Department of History, Temple University, Philadelphia, Pennsylvania, USA.
Stewart Wolpin, New York, New York, USA.
William E. Worthington, Jr., National Museum of American History, Smithsonian Institution, Washington, D.C., USA.
Jeffrey C. Wynn, U.S. Geological Survey, Reston, Virginia, USA.
Yeang, Chen Pang, Program in Science, Technology and Society, Massachusetts Institute of Technology, USA.
Ronald Young, History Department, Georgia Southern University, USA.
Thomas Zeller, Department of History, University of Maryland, USA.
Stelios Zimeras, Medcom GmbH, Germany.

A

Absorbent Materials

For thousands of years, plant-derived materials have served as the primary ingredient of absorbent materials. Jute, flax, silk, hemp, potatoes, and primarily cotton, have been employed since pre-Roman times. These simple plant-based fibers demonstrated molecular properties such as surface tension and colloid attraction, but it wasn't until the development of the ultramicroscope in 1903 that the size and structure of molecules was better understood and the actual chemical process of absorption grasped. The late nineteenth century inspired a new wave of design for the specialized applications of absorbent material—as sanitary napkins and diapers—and eventually helped drive innovative applications for the burgeoning fields of organic and polymer science in the twentieth century.

The need for sterile bandages in medicine precipitated the design of mass-producible, absorbent materials. In 1886, the medical supply company Johnson & Johnson developed surgical wound dressings made of heated, sterilized absorbent cotton with a gauze overlay to prevent fibers sticking to wounds. This design for sterile wound dressing became a fixed part of medical treatment, although it was still unavailable to the general public. However, as women changed their clothing styles and became more independent, demand increased for transportable absorbent menstrual napkins, as well as disposable diapers. In 1887 an American, Maria Allen, created a cotton textile diaper covered with a perforated layer of paper, to draw blood away from the skin, with a gauze layer stitched around it. It was an improvement over the usual washable cotton "rag" that was extremely leaky (as both a sanitary napkin and a diaper). However, it was too expensive for mass production.

Johnson & Johnson continued to improve on the absorption capacity of their original bandage. They discovered that heating and compressing several layers of cotton together provided higher absorption, less leakage, and less bulk in their dressings. When the Lister Towel, as it was named, became widespread in 1896, menstrual products such as the German-manufactured Hartman's Pads and bolts of "sanitary" cotton cloth appeared in catalogs for women. However, the Johnson & Johnson product was expensive. Cotton, while readily available, still had to be hand picked, processed and sterilized. So, in 1915, an American paper supply company called Kimberly–Clark developed Cellucotton, a bandage material that combined sterile cotton with wood pulp-derived cellulose. During World War I, nurses working in Europe began to use both the Lister Towel and Cellucotton as menstrual pads. By 1921, propelled by this innovative application, Kimberly–Clark manufactured Cellucotton-based disposable pads called Kotex. Thick, with a gauze overlay, they employed several different securing devices. Used in diapers, Cellucotton was sometimes covered by a thick rubber pant, which inhibited evaporation and could exacerbate diaper rash and urinary tract infections in babies. "Breathability" would become one of the challenges in the decades to come.

After the turn of the twentieth century, the molecular properties of most fibers were thoroughly understood. Protein fiber-based materials, such as wool, are made up of long, parallel, molecular chains connected by cross-linkages.

While able to absorb 30 percent of its weight, it would also expel liquid readily when squeezed, making it an unattractive menstrual or diaper material. Plant-based material such as cotton was made up of long chains of cellulose molecules arranged in a collapsed tube-like fiber. Cotton could easily absorb water by holding the water molecules within the tubes and between the fibers. In addition, the shape of the fibers meant that cotton could be easily manipulated by surfactants and additives. The rate of absorption depended largely on the surface tension between the absorbent material, and the fluid it was absorbing. Manipulating surface tension would become an element of future products.

For the first half of the twentieth century, absorbent materials varied little, but design changed dramatically. Tampons, available for millennia, now incorporated the new cotton-hybrid materials and by 1930 appeared widely on the market. In 1936, Dr. Earle C. Haas, an American physician, created and earned a patent for a cardboard tampon applicator. Soon thereafter, his product became the first Tampax brand tampon and was sold by Tambrands.

By 1938, American chemist Wallace Hume Carothers of the DuPont Company had helped create nylon, the first polymer textile, and it was soon included as a barrier to prevent leakage. In 1950, American housewife Marion Donovan created a plastic envelope from a nylon shower curtain that was perforated on one side and filled with absorbent cotton gauze. By 1973, scientists working at the Illinois-based National Center for Agricultural Utilization Research invented H-Span. They combined synthetic chemicals with cornstarch to create a uniquely absorbent polymer of hydrolyzed starch called polyacrylonitrile. The "Super Slurper," as it became known, was capable of absorbing up to 5,000 times its weight in water. In a dry powdered state, the polymer chains are coiled and then treated with carboxylate to initiate a faster colloid transfer of water molecules to the starch.

Soon afterwards, "superthirsty" fibers appeared in absorbent products around the world. By the late 1970s, disposable diapers included a layer of some sort of highly absorbent fibers, covered with a lightweight plastic or nylon shell that allowed for more evaporation without leakage. The American-based company Procter & Gamble introduced a "superthirsty" synthetic material, made up of carboxymethylcellulose and polyester, into their tampons. The product, named Rely, far surpassed the absorbency of other competing tampons. Under competitive pressure, Tambrands and Playtex both produced versions of superthirsty tampons using derivatives of polyacrylate fibers. Diaper designs began to include convenience features such as refastenable tapes, elastic legs, barrier leg cuffs, elasticized waistbands, and "fit" guides to guarantee less leakage. The popular creped cotton tissue interior was replaced with denser cellulose-fiber mats, utilizing a highly absorbent cotton treated with a surfactant to encourage rapid absorption by increasing the surface tension between water molecules and cotton.

Research continued and resulted in a new wave of polymer-manipulated superabsorbers, namely hydrophilic cross-linked polymers. Incorporating a three-dimensional polymeric structure, this material did not dissolve in water and could absorb in three dimensions. By 1980, Japanese scientists created the first disposable diaper incorporating a superabsorbent polymer. Procter & Gamble soon developed "ultra thin" pads using a crystalline polymer layer that would gel when it absorbed water. This design also included a "Dri-Weave" top sheet, separating the wearer from the absorbent layer and using a capillary-like, nonwoven material to inhibit a reverse flow.

In the late 1970s, a dramatic increase in cases of toxic shock syndrome appeared among users of superabsorbent tampons. Eventually, the "superthirsty" absorbent was found to encourage growth of the bacteria *Staphyloccocus aureus*. In the early 1980s more health problems seemed to be linked to improvements in absorption, and by 1986 Tambrands and Playtex had removed their polyacrylate tampons from the market. Six years later the U.S. Food and Drug Administration reported that trace amounts of dioxin used to bleach and sterilize cotton components of pads, tampons, and diapers could cause birth defects and possibly cancer.

At the beginning of the twenty-first century, pads were comprised of anything from an absorbent, compressed cotton and cellulose-pulp core, a plastic moisture-proof liner, a soft nonwoven textile for drawing moisture away from the skin (like viscose rayon and cotton blend), and chemicals such as polyacrylates to prevent leakage and keep the product from falling apart. Scientists working for the U.S. Department of Agriculture had discovered that the cellulose properties of ground chicken feathers could be manipulated and used as an absorbent material, utilizing billions of tons of discarded poultry-plant waste. The fibers are straight polymer chains—like cotton—making them highly absorbent. Internationally, the use of

tampons, disposable diapers, and sanitary napkins is still largely reserved for developed countries. However, as more innovative techniques reduce the reliance on expensive imported products (e.g., bird feathers), the convenience of absorbent technology may stretch beyond current economic, cultural, and geographic borders.

See also **Fibers; Synthetic; Semi-Synthetic**

LOLLY MERRELL

Further Reading

Asimov, I. *New Guide to Science*. New York, 1994, 533–550.

Gutcho, M. *Tampons and Other Catamenial Receptors*. Noyes Data Corporation, Park Ridge, NJ, 1979.

Hall, A. *Cotton-Cellulose: Its Chemistry and Technology*. E. Benn, London, 1924.

Park, S. The Modern and Postmodern Marketing of Menstrual Products. *J. Pop. Cult*., 30, 149, 1996.

Swasy, A. *Soap Opera*. Times Books, New York, 1993.

Useful Websites

U.S. Environmental Protection Agency fact sheet on acrylamide: http://www.epa.gov/ttn/uatw/hlthef/acrylami.html

Activated Carbon

Activated carbon is made from any substance with a high carbon content, and activation refers to the development of the property of adsorption. Activated carbon is important in purification processes, in which molecules of various contaminants are concentrated on and adhere to the solid surface of the carbon. Through physical adsorption, activated carbon removes taste and odor-causing organic compounds, volatile organic compounds, and many organic compounds that do not undergo biological degradation from the atmosphere and from water, including potable supplies, process streams, and waste streams. The action can be compared to precipitation. Activated carbon is generally nonpolar, and because of this it adsorbs other nonpolar, mainly organic, substances. Extensive porosity (pore volume) and large available internal surface area of the pores are responsible for adsorption.

Processes used to produce activated carbons with defined properties became available only after 1900. Steam activation was patented by R. von Ostreijko in Britain, France, Germany, and the U.S. from 1900 to 1903. When made from wood, the activated carbon product was called Eponite (1909); when made from peat, it was called Norit (1911). Activated carbon processes began in Holland, Germany, and the U.S., and the products were in all cases a powdered form of activated carbon mainly used for decolorizing sugar solutions. This remained an important use, requiring some 1800 tons each year, into the twenty-first century.

In the U.S., coconut char activated by steam was developed for use in gas masks during World War I. The advantage of using coconut shell was that it was a waste product that could be converted to charcoal in primitive kilns at little cost. By 1923, activated carbon was available from black ash, paper pulp waste residue, and lignite. In 1919, the U.S. Public Health Service conducted experiments on filtration of surface water contaminated with industrial waste through activated carbon. At first, cost considerations militated against the widespread use of activated carbon for water treatment. It was employed at some British works before 1930, and at Hackensack in New Jersey. From that time there was an interest in the application of granular activated carbon in water treatment, and its subsequent use for this purpose grew rapidly. As improved forms became available, activated carbon often replaced sand in water treatment where potable supplies were required.

Coal-based processes for high-grade adsorbent required for use in gas masks originally involved prior pulverization and briquetting under pressure, followed by carbonization, and activation. The process was simplified after 1933 when the British Fuel Research Station in East Greenwich, at the request of the Chemical Research Defence Establishment, began experiments on direct production from coke activated by steam at elevated temperatures. In 1940, Pittsburgh Coke & Iron Company, developed a process for producing granular activated carbon from bituminous coal for use in military gas masks. During World War II, this replaced the coconut char previously obtained from India and the Philippines. The large surface area created by the pores and its mechanical hardness made this new material particularly useful in continuous decolorization processes. The Pittsburgh processes developed by the Pittsburgh Activated Carbon Company were acquired in 1965 by the Calgon Company. In late twentieth century processes, carbon was crushed, mixed with binder, sized and processed in low-temperature bakers, and subjected to high temperatures in furnaces where the pore structure of the carbon is developed. The activation process can be adjusted to create pores of the required size for a particular application. Activation normally takes

place at 800–900°C with steam or carbon dioxide. Powdered activated carbon is suitable for liquid and flue gas applications—the granulated form for the liquid and gas phases, and pelleted activated carbon for the gas phase. Granulated activated carbon is used as a filter medium for contaminated water or air, while the powdered form is mixed into wastewater where it adsorbs the contaminants and is later filtered or settled from the mixture. Activated carbon has also been used in chemical analysis for prior removal and concentration of contaminants in water. Trade names for activated carbon used in these processes are Nuchar and Darco.

Activated carbon has been used in the large-scale treatment of liquid waste, of which the effluent from the synthetic dye industry is a good example. Synthetic dye manufacture involves reactions of aromatic chemicals, and the reactants and products are sometimes toxic. In addition to an unpleasant taste and odor imparted to water, this waste is also highly colored, complex, and invariably very difficult to degrade. Fortunately, many of the refractory aromatic compounds are nonpolar, the property that permits adsorption onto activated carbon. In the 1970s, three large dye-making works in New Jersey used activated carbon to remove aromatics and even trace metals such as toxic lead and cadmium from liquid waste. In two cases, powdered activated carbon was added to the activated sludge treatment process to enhance removal of contaminants. In a third case, following biological treatment, the liquid effluent was adsorbed during upward passage in towers packed with granular activated carbon. The spent carbon from this continuous process was regenerated in a furnace, and at the same time the adsorbed waste solute was destroyed.

In 1962, Calgon utilized activated granular carbon for treating drinking water, and at the end of the twentieth century, municipal water purification had become the largest market for activated carbon. The older methods that involved disposal of spent carbon after use were replaced by the continuous processes using granulated activated carbon. By continuous reuse of the regenerated activated carbon, the process is ecologically more desirable. Apart from the inability to remove soluble contaminants (since they are polar) and the need for low concentrations of both organic and inorganic contaminants, the cost of the carbon is the greatest limitation in the continuous process.

Activated carbon also found wide application in the pharmaceutical, alcoholic beverage, and electroplating industries; in the removal of pesticides and waste of pesticide manufacture; for treatment of wastewater from petroleum refineries and textile factories; and for remediation of polluted groundwater. Although activated carbons are manufactured for specific uses, it is difficult to characterize them quantitatively. As a result, laboratory trials and pilot plant experiments on a specific waste type normally precede installation of activated carbon facilities.

See also **Green Chemistry; Technology, Society, and the Environment**

ANTHONY S. TRAVIS

Further Reading

Cheremisinoff, P.N. and Ellersbusch, E. *Carbon Adsorption Handbook*. Ann Arbor, MI, 1978.

Hassler, J.W. *Activated Carbon*. Chemical Publishing, New York, 1963. Includes a historical survey on pp. 1–14.

Useful Websites

A Brief History of Activated Carbon and a Summary of its Uses: http://www.cee.vt.edu/program_areas/environmental/teach/gwprimer/group23/achistory.html

Calgon Carbon, Company History: http://www.calgoncarbon.com/calgon/calgonhistory.html

Chemviron Carbon: http://www.chemvironcarbon.com/activity/what/history/menu.htm

Adhesives

Adhesives have been used for about six millennia, but it was only from the first decade of the twentieth century that any significant development took place, with the introduction of synthetic materials to augment earlier natural materials. The driving force for development has been the needs of particular industries rather than technological advances themselves. The introduction of synthetic resins began in about 1909, but although the growth in plywood manufacture was accelerated by World War I, little innovation was involved. Significant advances began with World War II and the development of epoxy and urea/formaldehyde adhesives for the construction of wooden aircraft, followed by the phenol/formaldehyde/polyvinyl formal adhesives for bonding aluminum, which cannot generally be welded. Later, adhesive bonding in conjunction with riveting was applied to automobile construction, initially to high-performance models but increasingly to mass-produced vehicles. The fastening of composite materials is, with few exceptions, accomplished by use of adhesives.

If the forces of adhesion are to be effective, intimate contact must be established between two components, one of them a liquid that will wet and flow across the other before solidifying so that the bond can resist and transmit any applied force. This change of phase from liquid to solid is achieved in a variety of ways.

Solution-Based Adhesives

The earliest adhesives were all natural products such as starch and animal protein solutions in water. These are still in use for applications where only low strength is required (e.g., woodworking or attaching paper and similar materials).In these cases, the cost has to be low because the uses are high volume. Until about 1930 these were the main adhesives used in all carpentry and furniture. Polyvinyl acetate adhesives are now probably the most important range of water-based adhesives. The base polymer is dispersed in water to give an emulsion that has to be stabilized, usually with approximately 5 percent polyvinyl alcohol.

Solutions in organic solvents were first introduced in 1928, and they are now perhaps the most widely used adhesives both for manufacturing and for do-it-yourself purposes. Based on solutions of polychloroprene as base polymer dissolved in organic solvents, they provide a fairly strong "quick-stick" bond. Particular grades are extensively used in the footwear industry. Because of the toxic, environmentally unfavorable properties of the solvents, considerable efforts are being devoted to replacing these with water-based products, but these have not yet been entirely satisfactory.

Hot-Melt Adhesives

One of the oldest types of adhesive is sealing wax. Since about 1960, these hot-melt adhesives have been introduced initially for large-scale industrial use and more recently for small-scale and do-it-yourself uses. Polyethylene is extensively used as the base for hot-melt adhesives since it is widely available in a range of grades and at low cost. Ethylene vinyl acetate is similarly a useful base, and the two are commonly used in combination to give effective adhesives with application temperatures in the range of 160–190°C. This means that the adhesives have an upper limit of service use of perhaps 140°C, and the materials being joined must be able to withstand the higher temperature. These adhesives are quite widely used in large-scale manufacturing. However there are a considerable number of applications where the temperature involved for normal hot-melt adhesives is excessive. Consequently, in the 1990s a group of special formulations evolved that have an application temperature in the range of 90 to 120°C without any loss of adhesive strength. The most recent developments are adhesives that are applied as hot-melts and are then "cured" by various means. They have all the advantages of ease of application and quick achievement of useful strength supplemented by a much higher service temperature. Curing may be achieved either by heating to a higher temperature than that of application or by irradiation with an electron beam.

Reactive Adhesives

Reactive adhesives include epoxides, urethanes, phenolics, silicones, and acrylates.

Epoxides. Introduced in the early 1940s, these depend on three-membered epoxy or oxirane rings at the end of carbon chains with pendant hydroxyl groups, all of which react with various second components to produce thermoset polymers. The second components are principally amines or acid anhydrides. Generally the epoxides give bonds of considerable strength and durability, but until recently they tended to be too brittle for many purposes. Developments beginning in the 1970s have enhanced the toughness of these and other structural adhesives.

Urethanes. These involve the reaction of an isocyanate with an organic compound containing a hydroxyl group. Like the epoxides, variation of the properties of the final polymer can readily be controlled with two ingredients to give a product that may be an elastomer, a foam, or one that is stiff and bristle-like. Urethanes are increasingly used in a wide variety of situations.

Phenolics. The phenolics group of adhesives includes two that are somewhat different in their uses. The first, urea/formaldehyde formulations, were developed in the 1920s and 1930s and are mainly significant in the manufacture of plywood and similar products. The second group is phenol/polyvinyl formal formulations mainly used in aircraft construction for bonding aluminum and developed during World War II. Phenolics all involve curing under considerable pressure at an elevated temperature, typically 1500°C for 30 minutes at a pressure of 10 atmospheres for an aircraft adhesive. The bonds are of considerable strength and durability, suitable for primary aircraft structures.

Silicones. Silicones, generally silicone (or siloxane) rubbers, are largely used as sealants that combine adhesion with their gap-filling characteristics. Commonly used for sealing around baths and similar fittings, they cure by reaction with moisture from the environment. Industrially, particularly in automobile construction, there are many situations where providing a bond of moderate strength together with filling a gap between parts, which may amount to several millimeters, is required.

Acrylates. The acrylates include four types of adhesives.

1. Anaerobic adhesives (c. 1950) are formulations in which polymerization is prevented by the presence of oxygen. If oxygen is removed and ferrous ions are present, the liquid very quickly polymerizes to a hard, rather brittle solid. The main use for this is in thread locking in machinery and in the securing of coaxial joints.
2. Cyanoacrylates, or "super glues," were developed in 1957. They are colorless, very mobile liquids derived from cyanoacrylic acid. They readily polymerize, particularly in conjunction with the imperceptible film of moisture that is invariably present on surfaces. The bonds are very susceptible to attack by water and are only stable below about 80°C. Nevertheless, they are extensively used in product assembly in the electronics industry where they are likely to be exposed to only benign conditions.
3. Reactive acrylics (sometimes called "second generation" acrylates, developed in 1975) depend upon a polymerization reaction that follows a free radical path. This means that the ratio of the two components is relatively unimportant, so careful control of quantities is unnecessary. In parallel with the development of this system, a technique was perfected for increasing the toughness of the cured adhesive by incorporating minute particles of rubber. The adhesive is in two parts: a viscous gel and a mobile liquid. These two are spread one on each side of the joint. When the two are brought together, they react quickly to give a strong bond, which is handleable in 2 to 3 minutes, with full working strength in 1 hour and ultimate strength in 24 hours. These adhesives not only give a strong bond of high toughness very quickly, they are also able to contend with oily surfaces. They provide an exceedingly satisfactory product that meets a number of requirements in advanced assembly, especially within the automobile industry.
4. A series of acrylic adhesives has been produced which are cured by irradiation with ultraviolet light. Clearly they can only be used where the radiation can reach the adhesive; for example, where one component is transparent to the UV wavelength. While a considerable range of these products has been developed, very little information has been released about their composition.

High-Temperature Adhesives

All the adhesives considered so far can only provide useful bonds up to very limited temperatures, commonly 100°C or perhaps 150°C. There are demands, mainly military, for bonds that can withstand up to 300°C. To meet these needs, some adhesive base polymers have been developed that are based on carbon and nitrogen ring systems with a limited service life at these high temperatures.

Pressure-Sensitive Adhesives

Pressure-sensitive adhesives (e.g., Scotch Tape, first sold in 1940) are totally different from any others. These adhesives depend on an exceedingly high-viscosity liquid that retains this state throughout its life and never cross-links or cures. The strength of the bond is dependent on the pressure applied to it as the bond is made. The useful life of pressure-sensitive adhesives is generally limited to perhaps one or two years.

KEITH W. ALLEN

Further Reading

Adams, R.D., Comyn, J. and Wake, W.C. *Structural Adhesive Joints in Engineering,* 2nd edn. Chapman & Hall, London, 1997.

Comyn, J. *Adhesion Science*. Royal Society of Chemistry, Cambridge, 1997.

Kinloch, A.J. *Adhesion and Adhesives: Science and Technology*. Chapman & Hall, London, 1987.

Packham, D.E. *Handbook of Adhesion,* Longman, Harlow, 1992.

Pocius, A.V. *Adhesion and Adhesives Technology: An Introduction*, 2nd edn. Carl Hanser, Munich, 2002.

Wake, W.C. *Adhesion and the Formulation of Adhesives,* 2nd edn. Applied Science Publishers, London, 1982.

Agriculture and Food

In late-twentieth century Western societies, food was available in abundance. Shops and supermarkets offered a wide choice in products and brands. The fast-food industry had outlets in every neighborhood and village. For those in search of something more exclusive, there were smart restaurants and classy catering services. People chose what they ate and drank with little awareness of the sources or processes involved as long as the food was tasty, nutritious, safe, and sufficient for everyone. These conditions have not always been met over the last century when food shortages caused by economic crises, drought, or armed conflicts and war, occurred in various places. During the second half of the twentieth century, food deficiency was a feature of countries outside the Western world, especially in Africa. The twentieth century also witnessed a different sort of food crisis in the form of a widespread concern over the quality and safety of food that mainly resulted from major changes in production processes, products, composition, or preferences. Technology plays a key role in both types of crises, as both cause and cure, and it is the character of technological development in food and agriculture that will be discussed. The first section examines the roots of technological developments of modern times. The second is an overview of three patterns of agricultural technology. The final two sections cover developments according to geographical differences.

Before we can assess technological developments in agriculture and food, we must define the terms and concepts. A very broad description of agriculture is the manipulation of plants and animals in a way that is functional to a wide range of societal needs. Manipulation hints at technology in a broad sense; covering knowledge, skills, and tools applied for production and consumption of (parts or extractions of) plants and animals. Societal needs include the basic human need for food. Many agricultural products are food products or end up as such. However, crops such as rubber or flax and animals raised for their skin are only a few examples of agricultural products that do not end up in the food chain. Conversely, not all food stems from agricultural production. Some food is collected directly from natural sources, like fish, and there are borderline cases such as beekeeping. Some food products and many food ingredients are artificially made through complicated biochemical processes. This relates to a narrow segment of technology, namely science-based food technology.

Both broad and narrow descriptions of agriculture are relevant to consider. In sugar production for example, from the cultivation of cane or beets to the extraction of sugar crystals, both traditional and science-based technologies are applied. Moreover, chemical research and development resulted in sugar replacements such as saccharin and aspartame. Consequently, a randomly chosen soft drink might consist of only water, artificial sweeteners, artificial colorings and flavorings, and although no agriculture is needed to produce such products, there is still a relationship to it. One can imagine that a structural replacement of sugar by artificial sweeteners will affect world sugar prices and therewith the income of cane and beet sugar producers. Such global food chains exemplify the complex nature of technological development in food and agriculture.

The Roots of Technological Development

Science-based technologies were exceptional in agriculture until the mid-nineteenth century. Innovations in agriculture were developed and applied by the people cultivating the land, and the innovations related to the interaction between crops, soils, and cattle. Such innovation is exemplified by farmers in Northern Europe who confronted particular difficulties caused by the climate. Low temperatures meant slow decomposition of organic material, and the short growing season meant a limited production of organic material to be decomposed. Both factors resulted in slow recuperation of the soil's natural fertility after exploitation. The short growing season also meant that farmers had to produce enough for the entire year in less than a year. Farmers therefore developed systems in which cattle and other livestock played a pivotal role as manure producers for fertilizer. Changes in the feed crop could allow an increase in livestock, which produced more manure to be used for fertilizing the arable land, resulting in higher yields. Through the ages, farmers in Northern Europe intensified this cycle. From about the 1820s the purchase of external supplies increased the productivity of farming in the temperate zones. Technological improvements made increases in productivity not only possible but also attractive, as nearby markets grew and distant markets came within reach as a result of the nineteenth century transportation revolution.

An important development at mid-nineteenth century was the growing interest in applying

science to agricultural development. The two disciplines with the largest impact were chemistry and biology. The name attached to agricultural chemistry is Justus von Liebig, a German chemist who in the 1840s formulated a theory on the processes underlying soil fertility and plant growth. He propagated his organic chemistry as the key to the application of the right type and amount of fertilizer. Liebig launched his ideas at a time when farmers were organizing themselves based on a common interest in cheap supplies. The synergy of these developments resulted in the creation of many laboratories for experimentation with these products, primarily fertilizers. During the second half of the nineteenth century, agricultural experiment stations were opened all over Europe and North America.

Sometime later, experimental biology became entangled with agriculture. Inspired by the ideas of the British naturalist Charles Darwin, biologists became interested in the reproduction and growth of agricultural crops and animals. Botany and, to a lesser extent, zoology became important disciplines at the experimental stations or provided reasons to create new research laboratories. Research into the reproductive systems of different species, investigating patterns of inheritance and growth of plant and animal species, and experimentation in cross-breeding and selection by farmers and scientists together lay the foundations of genetic modification techniques in the twentieth century.

By the turn of the century, about 600 agricultural experiment stations were spread around the Western world, often operating in conjunction with universities or agricultural schools. Moreover, technologies that were not specifically developed for agriculture and food had a clear impact on the sector. Large ocean-going steamships, telegraphy, railways, and refrigeration, reduced time and increased loads between farms and markets. Key trade routes brought supplies of grain and other products to Europe from North America and the British dominions, resulting in a severe economic crisis in the 1880s for European agriculture. Heat and power from steam engines industrialized food production by taking over farm activities like cheese making or by expanding and intensifying existing industrial production such as sugar extraction. The development of synthetic dyes made crop-based colorants redundant, strongly reducing or even eliminating cultivation of the herb madder or indigo plants. These developments formed the basis of major technological changes in agriculture and food through the twentieth century.

Patterns of Technology Development

The twentieth century brought an enormous amount of technology developed for and applied to agriculture. These developments may be examined by highlighting the patterns of technology in three areas—infrastructure, public sector, and commercial factory—as if they were seen in cross section. The patterns are based on combined material and institutional forces that shaped technology.

A major development related to infrastructure concerns mechanization and transport. The combustion engine had a significant effect on agriculture and food. Not only did tractors replace animal and manual labor, but trucks and buses also connected farmers, traders, and markets. The development of cooling technology increased storage life and the distribution range for fresh products. Developments in packaging in general were very important. It was said that World War I would have been impossible without canned food. Storage and packaging is closely related to hygiene. Knowledge about sources and causes of decay and contamination initiated new methods of safe handling of food, affecting products and trade as well as initiating other innovations. In the dairy sector, for example, expanding markets led to the growth and mergers of dairy factories. That changed the logistics of milk collection, resulting in the development of on-farm storage tanks. These were mostly introduced together with compression and tube systems for machine milking, which increased milking capacity and improved hygiene conditions. A different area of infrastructure development is related to water management. Over the twentieth century, technologies for irrigation and drainage had implications for improved "carrying capacity" of the land, allowing the use of heavy machinery. Improved drainage also meant greater water discharge, which in turn required wider ditches and canals. Water control also had implications for shipping and for supplies of drinking water that required contractual arrangements between farmers, governing bodies, and other agencies.

During the twentieth century, most governments supported their agricultural and food sectors. The overall interest in food security and food safety moved governments to invest in technologies that increased productivity and maintained or improved quality. Public education and extension services informed farmers about the latest methods and techniques. Governments also became directly involved in technological development, most notably crop improvement. Seed is a difficult product to

exploit commercially. Farmers can easily put aside part of the harvest as seed for the next season. Public institutes for plant breeding were set up to improve food crops—primarily wheat, rice, and maize—and governments looked for ways to attract private investment in this area. Regulatory and control mechanisms were introduced to protect commercial seed production, multiplication, and trade. Private companies in turn looked for methods to make seed reproduction less attractive to farmers, and they were successful in the case of so-called hybrid maize. The genetic make-up of hybrid maize is such that seeds give very high yields in the first year but much less in the following years. To maintain productivity levels, farmers have to purchase new seed every season. Developments in genetic engineering increased the options for companies to commercially exploit seed production.

Most private companies that became involved in genetic engineering and plant breeding over the last three decades of the twentieth century started as chemical companies. Genetic engineering allowed for commercially attractive combinations of crops and chemicals. A classic example is the herbicide Roundup, developed by the chemical company Monsanto. Several crops, most prominently soy, are made resistant to the powerful chemical. Buying the resistant seed in combination with the chemical makes weed control an easy job for farmers. This type of commercial development of chemical technologies and products dominated the agricultural and food sector over the twentieth century. Artificially made nitrogen fertilizers are one such development that had a worldwide impact. In 1908, Fritz Haber, chemist at the Technische Hochschule in Karlsruhe, fixed nitrogen to hydrogen under high pressure in a laboratory setting. To exploit the process, Haber needed equipment and knowledge to deal with high pressures in a factory setting, and he approached the chemical company BASF. Haber and BASF engineer Carl Bosch built a crude version of a reactor, further developed by a range of specialists BASF assigned to the project. The result was a range of nitrogen fertilizer products made in a capital and knowledge-intensive factory environment. This type of development was also applied to creating chemicals such as DDT for control of various pests (dichloro-diphenyl-trichloroethane), developed in 1939 by Geigy researcher Paul Müller and his team. DDT may exemplify the reverse side of the generally positive large-scale application of chemicals in agricultural production—the unpredictable and detrimental effects on the environment and human health.

The commercial factory setting for technology development was omnipresent in the food sector. The combination of knowledge of chemical processes and mechanical engineering determined the introduction of entirely new products: artificial flavorings, products, and brands of products based on particular food combinations, or new processes such as drying and freezing, and storing and packaging methods.

Patterns of Technology Development in the Western World

Technological developments in agriculture and food differ with regard to geography and diverging social and economic factors. In regions with large stretches of relatively flat lands, where soil conditions are rather similar and population is low, a rise in productivity is best realized by technologies that work on the economies of scale. The introduction of mechanical technologies was most intensive in regions with these characteristics. Beginning early in the twentieth century, widespread mechanization was a common feature of Western agriculture, but it took different forms. In the Netherlands, for example, average farm size was relatively small and labor was not particularly scarce. Consequently, the use of tractors was limited for the first half of the twentieth century as emphasis was placed on improved cultivation methods. Tractors became widely used only after the 1950s when equipment became lighter and more cost-effective and labor costs rose sharply. The result was an overall increase of farm size in these regions as well. The Dutch government changed the countryside with a land policy of connecting and merging individual parcels as much as possible. This huge operation created favorable conditions for expansion; but where the land was already under cultivation, the only way to expand was to buy up neighboring farms. The effect was a considerable reduction in the number of farm units. An exception to this process was the Dutch greenhouse sector, in which improvements in construction, climate regulation, and introduction of hydroponic cultivation, increased production without considerable growth of land per farm unit.

The Dutch greenhouse sector is also an exemplary case of technological support in decision making and farm management. In Western countries a vast service sector emerged around agriculture and food. This process in fact started early in the twentieth century with the rise of extension services, set up as government agencies or private companies. Experimental methods

based on multivariate statistics, developed by the British mathematician Karl Fisher, are the major tool in turning results of field experiments into general advisories. In keeping with the development of modern computers, digital models of crop growth and farming systems became more effective. Computer programs help farmers perform certain actions and monitor other equipment and machinery; yet even in the most technologically advanced greenhouses, the skilled eye of the farmer is a factor that makes a considerable difference in the quality and quantity of the final product.

The means by which agriculture in the West raised productivity have been questioned. Doubts about the safety of food products and worries over the restoration of nature's capacity became recurrent issues in public debate. Moreover, technological advances in tandem with subsidies resulted in overproduction, confronting national and international governing bodies with problems in trade and distribution, and a public resistance against intensive agriculture, sometimes called agribusiness. Technology is neither good nor bad; much of the knowledge underlying technologies with a detrimental effect also helps detect polluting factors and health hazards. Although a substantial part of research and technological efforts are aimed at replacing and avoiding harmful factors, many such "clean" technologies are commercially less interesting to farmers and companies. Subsidies and other financial arrangements are again being used to steer technology development, this time in the direction of environmentally friendly and safe forms of production.

Patterns of Technology Development in Less Developed Countries

From the beginning of the twentieth century, scientific and technological developments in the agricultural and food sector were introduced to less developed countries either by Western colonizing powers or by other forms of global interaction. The search for improved farming methods and new technology were mostly institutionalized at existing botanical gardens and established in previous centuries. Plant transfer and economic botany were a major modality of twentieth century technological improvement in less developed countries.

The early decades of the century featured an emphasis on technological improvement for plantation agriculture. Plantation owners invested in scientific research for agriculture, often supported by colonial administrations. The gradual abolition of slavery during the nineteenth century, increasing labor costs, was a reason to invest in technology. Other factors were more specific to particular sectors; for example, the rise of European beet sugar production encouraging cane sugar manufacturers to invest in technological improvement. Another example was the emergence of the automobile industry, which initiated a boom in rubber production.

Most colonial administrations launched programs, based on the combination of botanical and chemical research, to improve food crop production in the first decades of the twentieth century. It was recognized that dispersion of new technologies to a small number of plantation owners was different from initiating change among a vast group of local food crop producers. The major differences concerned the ecology of farming (crop patterns and soil conditions) and the socioeconomic conditions (organization of labor or available capital). Agronomists had to be familiar with local farming systems, occasionally resulting in pleas for a technology transfer that would better meet the complexity of local production. The overall approach, however, was an emphasis on improvement of fertilization and crop varieties. Transfer of the Western model gained momentum in the decades after World War II. Food shortages in the immediate postwar years encouraged European colonial powers to open up large tropical areas for mechanized farming. Unfortunately, the result was largely either a short-lived disaster, as in the case of the British-run groundnut scheme in Tanzania, or a more enduring problem, as in case of the Dutch-run mechanized rice-farming schemes in Surinam. The 1940s also saw the beginnings of a movement that came to be known as the "green revolution." Driven by the idea that hunger is a breeding ground for communism, American agencies initiated a research program for crop improvement, primarily by breeding fertilizer-responsive varieties of wheat and rice. Agencies were put together in a Consultative Group on International Agricultural Research (CGIAR). Technological progress was realized by bringing together experts and plant material from various parts of the world. Modified breeding techniques and a wide availability of parent material resulted in high-yielding varieties of wheat and rice. Encouraged by lucrative credit facilities, farmers, especially in Asia, quickly adopted the new varieties and the required chemicals for fertilization and pest control. Research on the adoption process of these

varieties made clear that many farmers modified the seed technology based on specific conditions of the farming systems. In areas where such modifications could not be achieved—primarily rice growing regions in Africa—green revolution varieties were not very successful. Based on these findings, CGIAR researchers began to readdress issues of variation in ecology and farming systems. This type of research is very similar to that done by colonial experts several decades earlier. However, because of decolonization and anti-imperialist sentiments among Western nations, much of this earlier expertise has been neglected. This is just one of the opportunities for further research in the domain of agriculture and food technology.

See also **Biotechnology; Breeding: Animal, Genetic Methods; Breeding: Plant, Genetic Methods; Dairy Farming; Farming, Agricultural Methods; Fertilizers; Food Preservation; Irrigation; Pesticides; Transport, Foodstuffs**

HARRO MAAT

Further Reading

Baum, W.C. and Lejeune, M.L. *Partners Against Hunger: The Consultative Group on International Agricultural Research*. World Bank, Washington, 1986.

Beckett, J.V. *The Agricultural Revolution*. Blackwell, Oxford, 1990.

Bray, F. *The Rice Economies: Technology and Development in Asian Societies*. Blackwell, Oxford, 1986.

Brockway, L.H. *Science and Colonial Expansion: the Role of the British Royal Botanic Gardens*. Academic Press, New York, 1979.

Bonano, A., Busch L. and Friedland W.H., Eds. *From Columbus to ConAgra: the Globalization of Agriculture and Food*. University Press of Kansas, Lawrence, Kansas, 1994.

Busch, L., Ed. *Science and Agricultural Development*. Allenheld Osmun, New Jersey, 1981.

Cittadino, E. *Nature as the Laboratory: Darwinian Plant Ecology in the German Empire, 1800–1900*. Cambridge University Press, Cambridge, 1990.

Fitzgerald, D. *The Business of Breeding: Hybrid Corn in Illinois, 1890–1940*. Cornell University Press, London, 1990.

Gigerenzer, G. *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge University Press, Cambridge, 1989.

Hurt, R. D. and M.E. Hurt. *The History of Agricultural Science and Technology: An International Annotated Bibliography*. Garland, New York, 1994.

Headrick, D.R. *The Tentacles of Progress: Technology Transfer in the Age of Imperialism, 1850–1940*. Oxford University Press, Oxford, 1988.

Kimmelman, B.A. *A Progressive Era Disciplie: Genetics at American Agricultural Colleges and Experiment Stations, 1900–1920*. UMI, Ann Arbor, MI, 1987.

Kiple, K.F. and Ornelas, K.C., Eds. *The Cambridge World History of Food*. Cambridge University Press, Cambridge, 2000.

Maat, H. *Science Cultivating Practice: A History of Agricultural Science in the Netherlands and its Colonies*. Kluwer, Dordrecht, 2001.

Perkens, J.H. *Geopolitics and the Green Revolution: Wheat, Genes and the Cold War*. Oxford University Press, New York, 1997.

Richards, P. *Indigenous Agricultural Revolution: Ecology and Food Production in West Africa*. Unwin Hyman, London, 1985.

Rossiter, M.W. *The Emergence of Agricultural Science: Justus Liebig and the Americans, 1840–1880*. Yale University Press, New Haven, 1975.

Slicher van Bath, B.H. *The Agrarian History of Western Europe A.D. 500–1850*. Edward Arnold, London, 1963.

Zanden, J.L. *The Transformation of European Agriculture in the Nineteenth Century: the Case of the Netherlands*. Free University Press, Amsterdam, 1994.

Air Conditioning

Air conditioning is a mechanical means of controlling the interior environment of buildings through the circulation, filtration, refrigeration and dehumidification of air. Although commonly thought of as a method of *cooling* interiors, it treats the interrelated factors of temperature, humidity, purity, and movement of air and is closely linked to developments in building electrification. The history of the modification of building air by human beings goes back millennia, and the major components of what we now call air conditioning, such as forced ventilation and mechanical refrigeration, were well developed by the nineteenth century. Fifty years before the term air conditioning was used to describe a mechanical system, the new Houses of Parliament in London were built to David Boswell Reid's comprehensive and controversial scheme for cooling, heating, purifying, humidifying, and circulating air. Outside air was drawn into a duct, passed through filtering sheets, heated by a hot water heater or cooled over blocks of ice, then drawn up through holes in the floor into the House of Commons. The used air was drawn up by the heat of a ventilating fire through raised panels in a glass ceiling and discharged.

The term "air conditioning" was introduced as a technical term in the twentieth century by an American textile engineer, Stuart W. Cramer. He used it to describe a process by which textile mills could be humidified and ventilated so as to "condition" the yarn produced there. Both his wall and ceiling units drew air into their casings, where it was pulled through a water spray and a cloth filter and discharged. Cramer's hygrometer

measured the heat and relative humidity of the air and controlled the entire system automatically.

Cramer's contemporary Willis Carrier was among the most relentless researchers and promoters of "man-made weather," and while he is known as the father of air conditioning, the range of processes and products involved in air conditioning cannot be attributed to a single author. Carrier's "Apparatus for Treating Air" was only used to cleanse air that was already circulating through existing ventilation systems. Technologies that would culminate in room cooling were being developed in the refrigeration industry contemporaneously with Carrier's work in humidification. For centuries, ice and water had been manipulated to cool air that circulated in theaters, hospitals, factories, and other large public spaces. The New York Stock Exchange was air-cooled in 1904 using a toxic ammonia refrigerant. Air was channeled from roof level to the basement where it was filtered through layers of hanging cheesecloth and brine-chilled coils, and then blown into the building by a fan.

The development of industrial air conditioning, also called process air conditioning, dominated the newly created industry at the beginning of the twentieth century. Each air conditioning system was custom designed by engineers for the buildings into which they were installed. Human comfort was a byproduct of technologies aimed at improving industrial production. Comfort began to be promoted as a luxury in the 1920s when thousands of "evaporating–cooling stations" were installed in movie theaters where they relieved crowds of heat and smells. The J.L. Hudson Department Store in Detroit, Michigan, was the first commercial space to benefit from a "centrifugal chiller" installed by the Carrier Corporation in 1924.

Although Alfred Wolff had installed air conditioning systems in elite houses in the last decade of the nineteenth century, significant challenges remained in the twentieth century to the mass production of domestic room or central air conditioners: size, weight, cost, safety, and limitations on electrical service capacity for domestic buildings. One of the earliest room coolers, the Frigidaire Room Cooler of 1928, weighed 270 kilograms, required 2.7 kilograms of a toxic refrigerant, and was available in a single output size of 5500 Btu per hour. A room cooler of this size could only be (and often was) used in violation of electrical codes. Early room coolers cost in the thousands of dollars. It is not surprising that in 1930, General Electric could sell and install only thirty of the DH-5 models, the casing of which resembled a Victrola phonograph cabinet. Air conditioning began to be marketed as a comfort device for domestic consumption during the 1930s as manufacturers and mass production techniques helped democratize a product that was expensive, cumbersome, and custom designed and installed. The deprivation of the Great Depression followed by the post-World War II housing boom (particularly in the United States) facilitated the mass production and installation of domestic air conditioning across the class spectrum.

The development of chlorofluorocarbon gas (CFC or Freon) by Thomas Midgely in the 1920s, its manufacture by DuPont, and its successful application in the refrigeration industry galvanized the development of air conditioning as both a cooling device and mass-market product. Freon was considered the first nontoxic refrigerant and a replacement for sulfur dioxide, carbon dioxide, and ammonia. Together with development of the hermetically sealed motor compressor and lightweight finned coils it provided the foundation for air conditioning in its current form.

Air conditioners manufactured after the early 1930s, whether placed in a window or installed as part of the internal ducting of a building (central air conditioning), have included five basic mechanical components: compressor, fan (often two), condenser coil, evaporator coil, and chemical refrigerant. In central, or "split," systems there is a hot side, located outside the building, and a cold side, located inside. On the cold side, hot indoor air is drawn into the furnace and blown over an evaporator coil in which the refrigerant liquid is located. The refrigerant absorbs the heat of the air and evaporates, and the cooled air is then circulated throughout the building in internal ducts. The evaporated refrigerant is then pumped through a compressor and over condenser coils in contact with outside air. Once the heat of the refrigerant is transferred to the outside air, it liquefies and recirculates through an expansion valve and back into the cold side of the system. The hot air is expelled to the outside by a fan. A window air conditioner, which is powered by electricity and can weigh in the 25 kilograms range, also contains a hot and cold side, located on the outside and inside of the window, respectively. Other than the fact that the hot and cold sides of the central system are split and that the central system has a much higher capacity, these two systems function on essentially identical principles. Large buildings, taxed by extensive piping and larger capacities, often employ chilled water systems and cooling towers.

By 1970, over five million room air conditioners were being produced per year. By the end of the twentieth century, over 80 percent of single-family houses in the United States had air conditioning.

Since the 1970s, developments in air conditioning have focused on efficiency and environmental concerns. Freon was discovered to be destroying the ozone layer, and restrictions on its use and manufacture were imposed. It has been replaced in air conditioners by safer coolants such as hydrofluorocarbons and hydrochlorocarbons. A movement to replace thousands of Freon or CFC air conditioners with higher efficiency, non-CFC models was underway at the end of the twentieth century.

See also **Refrigeration, Mechanical**

JENNIFER COUSINEAU

Further Reading

Banham, R. *The Architecture of the Well Tempered Environment*. Architectural Press, London, 1969.
Bruegman, R. Central heating and forced ventilation: origins and effects on architectural design. *J. Soc. Archit. Hist.*, October 1978, 143–160.
Cooper, G. *Air-Conditioning America: Engineers and the Controlled Environment, 1900–1960*. Johns Hopkins University Press, Baltimore, 1998.
Elliot, C.D. *Technics and Architecture: The Development of Materials and Systems for Buildings*. MIT Press, Cambridge, MA, 1992.
Faust, F.H. The early development of self-contained and packaged air conditioners. *J. ASHRAE*, 1986, 353–369.
Fitch, J. M. and Bobenhausen, W. *American Building: The Environmental Forces That Shape It*. Oxford University Press, New York, 1999.
Ingels, M. *Father of Air Conditioning*. Country Life Press, Garden City, NY, 1952.

Air Traffic Control Systems

When a technology is fully assimilated, it is relatively invisible. Such is the way with air traffic control. It seems to go on all around us without our caring much about it, even if we fly.

Air traffic control began via the mail service. The airmail had to be delivered, and many new routes in strange places were pioneered by the need to fly mail. Mail pilots used writing on building roofs below to guide them during the day and various lights at night. However, most of the congestion was at the airports themselves. Controllers would be stationed in a tower or at the end of a runway and would use light signals (still the same ones used in a communications failure today) to guide landings or to hold aircraft for a break in the traffic.

When airplanes were rare, the threat of collision, even in cloudy weather, was low. As the number of commercial passenger airlines grew in the 1930s, a method of separating them from one another and a way of proceeding as safely in the clouds was necessary. Radio was not in widespread aerial use, and radar had not yet been invented. In this depression era, the U.S. government had little money, so it required the airlines to establish Air Traffic Control Units (ATCUs) to separate and space the traffic. Markers, called "shrimp boats," which could carry information about an airplane's intended altitude and air speed, represented an aircraft on a flight plan. The controllers moved the shrimp boats on maps following the approximate motion of the planes. In 1937, the federal government had enough money to incorporate the ATCUs. Units were renamed Air Traffic Control Stations (ATCS) and most controllers became civil service employees.

Cleveland had the first radio installed in a tower in the late 1930s, but there were few other developments until after World War II. The U.S. became crisscrossed by very high frequency (VHF) omnidirectional range transmitters (VORs), which were used for navigation. The rest of the world, including the U.K., which had many former bomber bases, used radio beacons that were much less accurate. With the VORs, it was possible to proceed on an exact path in the sky.

It was not until 1956 that the first radar dedicated to air traffic control was installed in Indianapolis, Indiana. This development was slow in coming because radar used for interceptions was different from radar used for spacing air traffic. Controllers still used the shrimp boats, but the radar could tell if a flight was progressing as planned.

Finally, computers came to be used. They would store the flight plans and drive displays that showed the controllers the identification, altitude, and airspeed of airplanes. Each aircraft was required to carry a transponder that could broadcast a four-digit code to enable flights to be paired with their radar track. Some of these devices had nearly a quarter-century of service when they were retired. The airspace system founded with their help enabled separation of only 300 meters vertically and 8 kilometers horizontally. The rest of the world based their air traffic control on the American system, with some differences where there is intermittent radar coverage, like the 80-kilometer separation in South America.

In the modern air traffic system, preparation begins when either the pilot or the dispatch office of an airline files a flight plan with a preferred route,

altitude, and airspeed. Before departing, the pilot calls Clearance Delivery to obtain a flight clearance, which may include changes in the route due to expected traffic or to weather problems, and which also includes a transponder identifier number. The pilot sets this number into the transponder, calls Ground Control for taxi clearance, and taxies the plane to the appropriate runway for take-off.

After arriving at the runway entrance and performing any last-minute checks, the pilot calls the tower air traffic controller. The controller tells the pilot when to enter the runway to avoid take-off and landing traffic and makes sure that there is enough separation to avoid invisible wake turbulence from airplanes that have already left. Cleared for take-off, the aircraft does its ground roll down the runway to rotation speed (the speed at which it can lift off), climbs, and retracts the landing gear.

Once the wheels have entered their wells, the tower tells the pilot to call Departure Control. This frequency was given as part of the clearance, which also gives an initial altitude. The departure controller has a flight strip for the plane printed out by a computer before the aircraft takes off. It contains the desired altitude. When traffic allows, the controller clears the plane to climb to that altitude.

Airplanes going west are at even-numbered altitudes, those going east at odd ones; thus the vertical separation is maintained by pilots' requests for appropriate altitudes. Horizontal separation is the controller's job, as well as monitoring vertical separation. For most of the flight, a Center controller is watching the airplane. The U.S. is divided into only 22 Centers, each with sectors assigned to controllers at displays with flight strips.

Approaching the destination airport, the process is essentially reversed. Within about 50 kilometers of the airport, the pilot speaks with Approach Control and is then handed off to the tower controller. When the controller gives the pilot the order "cleared to land," the field is the property of the pilot. At that point pilots can do anything they need to get their airplanes down. After landing, taxi instructions are given by Ground Control; at large airports, airplanes are transferred to a ramp controller for "parking."

Today's relatively close separation and ability to handle many thousands of flights make air traffic control one of the twentieth century's most ubiquitous technologies. It is also one of its most successful.

See also **Radar Aboard Aircraft; Civil Aircraft, Jet Driven; Civil Aircraft, Propeller Driven.**

JAMES E. TOMAYKO

Further Reading

National Research Council. *The Future of Air Traffic Control: Human Operators and Automation*, Wickens, C.D. and Parasuraman, R., Eds. National Academy Press, Washington D.C., 1998.

National Research Council. *Flight to the Future: Human Factors in Air Traffic Control*. Wickens, C.D., Mayor, A.S. and McGee J.P., Eds. National Academy Press, Washington D.C., 1997.

Nolan, M.S. *Fundamentals of Air Traffic Control*. Wadsworth, Belmont, CA, 1990.

Smolensky, M.W. and Stein, E.S., Eds. *Human Factors in Air Traffic Control*. Academic Press, San Diego, 1998.

Aircraft Carriers

Three nations built fleets of aircraft carriers—Britain, Japan and the United States—and each contributed to carrier design trends. Experiments began before World War I when, in November 1910, Eugene Ely flew a Curtiss biplane from a specially built forward deck of the cruiser *USS Birmingham* moored off Hampton Roads, Virginia. Two months later he accomplished the more difficult task of *landing* on a deck built over the stern of the cruiser *Pennsylvania*. Sandbags were used to anchor ropes stretched across the deck to help stop the airplane, which trailed a crude hook to catch the ropes.

Late in World War I, the British Royal Navy was the first to develop crude carriers when they adapted several merchant or naval vessels to carry seaplanes. The light cruiser *Furious* was converted in 1917 to carry eight aircraft, using a canted forward deck to allow them to take off; they then landed at sea or at nearby airstrips. By 1918 the ship's modified aft deck allowed separate take-offs and landings. *Furious* was thus the world's first true aircraft carrier. She was modified in the early 1920s to remove her bridge structure and extend the flying deck for nearly the length of the vessel, and she was rebuilt again in the 1930s. *Furious* and her two sister ships helped to pioneer the use of powered elevators (to raise or lower aircraft between the lower hangar and the landing and take-off deck) and of aircraft catapults, both of which would become standard carrier equipment.

In July 1917 the Royal Navy ordered the 10,000-ton *Hermes,* the world's first carrier designed as such. She set the design model of an "island" bridge on the right-hand side of the deck that was followed for years. In April 1942 she also became the first carrier lost to aircraft flown from another carrier. The 1938 *Ark Royal* was more than twice the size and was fitted with the enclosed "weather"

bow and stern that became hallmarks of British carriers, but it lacked the radar already being employed by American vessels. The postwar 37,000-ton *Eagle* and a second *Ark Royal* saw extensive modification over time, including the British-developed angle-deck extension allowing for simultaneous take-off and landing operations. The three ships of the 16,000-ton *Invincible* class, commissioned in the early 1980s, were soon fitted with "ski jump" bows to assist jet fighter take-off from their relatively short decks.

Given its Pacific location, the Japanese Imperial Navy was an early convert to the importance of aircraft carriers. In 1922 a converted oiler became *Hosho,* the country's first carrier, designed with British help. The 1927 *Akagi* was the country's first large (30,000 ton) fleet carrier, converted from an unfinished cruiser. At the time the Americans were making similar conversions. As with American (but not British) carriers of the time, the flight decks of Japanese vessels were made of wood to save weight and allow rapid repair when damaged. The best Japanese carriers were the two *Shokakus* of 1941. Displacing 26,000 tons and carrying up to 84 aircraft, they were more heavily armored and armed than their predecessors. The mid-1942 battles of the Coral Sea and Midway were the first naval confrontations where combat was conducted solely by carrier-based airplanes. In 1944 the 30,000-ton *Taiho* was the first Japanese carrier to feature an armored deck as well as enclosed bows for better sea keeping, as was common with British carriers of the period. Paralleling British and American wartime practice, many smaller "escort" carriers were created using military or merchant ship hulls. Japan constructed no aircraft carriers after 1945.

America's carrier experience began in 1922 with the small (13,000 ton) *Langley,* converted from a naval collier. She featured a compressed-air catapult to launch aircraft. The *Saratoga* and *Lexington* were converted from battle cruiser hulls and displaced nearly 40,000 tons when launched in 1927. Both vessels featured hydraulic aircraft elevators. The *Ranger* of 1934 was the first keel-up American carrier, though she proved to be too small for practical application. The three-ship *Yorktown* class launched between 1937 and 1941 featured light wooden decks, open bow and stern structures, multiple airplane elevators, and the ability to carry nearly 100 aircraft. No matter their size, however, carriers never had sufficient space. This led to such expedients as parking aircraft with only their front wheels on deck, and their tails hanging over the sea, supported by special extended bars. Biplane fighters were suspended from hangar ceilings to store them out of the way. More important was the folded wing, first developed by the British in the late 1930s, and adopted by the U.S. Navy early in 1941. Folding up one third or one half of each wing allowed many more aircraft on crowded decks and in hangar spaces.

American wartime carrier production featured the 24-ship *Essex* class of 27,000-ton vessels. The largest class of big carriers built by any nation, they could carry 91 aircraft. Ship antitorpedo protection was much improved, as were the elevators (more and better located) and anti-aircraft armament. Several of these ships were vastly modified in postwar years to take angle decks and jet aircraft. Completed only as the war ended, three vessels of the *Midway* class were, at more than 48,000 tons, the largest carriers to enter service until 1955. They were also the first American carriers to adopt the British-developed armored flight deck, which had saved several bombed British carriers during the war. They featured huge aviation gasoline storage capacity, large hangars, and enough space that they were substantially rebuilt to accommodate jet aircraft in the postwar years.

The *Forrestal* class of 62,000-ton carriers in the late 1950s became the basic model for all subsequent American vessels. Built for jet operations, the *Forrestals* featured an armored and canted deck to allow for simultaneous take-off and landing operations. Each of the four ships carried 90 aircraft and a crew of more than 4,000. Early in their service, anti-aircraft guns were removed and replaced with surface-to-air missile batteries that would figure on future carriers. Electronics and other features were updated continually. Four "improved" carriers of about the same size were added to the fleet in the 1960s with larger flight decks, improved electronics, and missile rather than gun defense.

The *Enterprise*, America's first atomic-powered carrier, entered service in 1961 with a range of 320,000 kilometers, or capable of four years' cruising. She was similar to the *Forrestal* carriers except for her small square island structure that originally featured "billboard" radar installations. Despite a huge cost increase (about 70 percent more than the *Forrestals*), she became the prototype for the ultimate *Nimitz* class of nuclear carriers that began to enter fleet service in the mid-1970s. Displacing nearly 95,000 tons, each had a crew of some 6,500 men. Driven by concerns about the growing expense of building and operating the huge American fleet carriers and their

vulnerability, research into smaller carrier designs continued.

See also **Battleships**

CHRISTOPHER H. STERLING

Further Reading

Blundell, W.G.D. *British Aircraft Carriers*. Model & Allied Publications, Hemel Hempstead, 1969.

Chesneau, R. *Aircraft Carriers of the World, 1914 to the Present: An Illustrated Encyclopedia*. Naval Institute Press, Annapolis, MD, 1984.

Friedman, N. *U.S. Aircraft Carriers: An Illustrated Design History*. Naval Institute Press, Annapolis, MD, 1983.

Musciano, W.A. *Warbirds of the Sea: A History of Aircraft Carriers and Carrier-Based Aircraft*. Schiffer, Atglen, PA, 1994.

Polmar, N. *Aircraft Carriers: A Graphic History of Carrier Aviation and Its Influence on World Events*. Doubleday, Garden City, NY, 1969.

Terzibaschitsch. S. *Aircraft Carriers of the U.S. Navy,* 2nd edn. Naval Institute Press, Annapolis, MD, 1989.

Aircraft Design

No innovation is more distinctive of the twentieth century, or more influential on its life or imagination, than the airplane. While the root technologies were established in the nineteenth century, it was the Wright brothers in 1903 who first synthesized them into a machine capable of sustained, controlled flight. Heavier-than-air flight stretched the limits of engine power and structural strength per unit weight and gave great impetus to technology, with influence far beyond aeronautics. It also largely initiated the engineering science of aerodynamics, again with far-reaching implications. The scope of aircraft design through the twentieth century can be considered by viewing the developments of four types of aircraft:

1. Biplane and externally braced wings
2. Streamlined monoplanes
3. Transonic aircraft
4. Supersonic and hypersonic aircraft.

Biplane and Externally Braced Wings

For their first four decades, successful airplanes were almost exclusively powered by spark-ignition internal combustion piston engines. Early engines followed automotive practice but gave greater attention to weight reduction and utilized lightweight materials such as aluminum for the block. Take-off power was typically about 40 kW, and weight per unit power was about 3 kilograms per kilowatt (kg/kW). Both overhaul life and mean time between in-flight failure were no more than a few hours. The theory of the piston engine was understood with reasonable clarity, but lack of experience and refinement in mechanical detail imposed severe limits.

The most distinctive feature of airplanes was the wing. Designers relied on avian models to frame their ideas for airfoil sections and on primitive wind tunnels, developed since the 1870s, to test them. A thin airfoil relative to its chord length was thought necessary for efficiency. To make a thin wing strong enough with the existing materials (generally wood, tension-braced with steel wire) external bracing was necessary. The Wrights joined two wings with a wood and wire truss structure in a biplane configuration, and most others followed them, although some monoplanes with external bracing were also seen as well as a few triplanes.

It had long been appreciated that an unstabilized wing would pitch down. This could be corrected by an aft-mounted horizontal tailplane surface rigged to provide negative lift, or a forward surface providing positive lift, the so-called "canard," as employed by the Wrights. Horizontal tailplanes were usually combined with a vertical stabilizer—a tail-aft configuration—because this lent itself to the most practical distribution of masses along the length of the aircraft. In particular, a tail-aft airplane could mount its engine forward for best cooling. A truss fuselage carried the tail at one end and the engine at the other, with the pilot accommodated amidships. Airfoil surfaces were covered with sewn and varnished fabric, and increasingly so was the fuselage. Other configurations were tried, but this one quickly became dominant for most applications because it offered the lowest weight and reasonable drag.

As the Wrights were the first to clearly recognize, control was crucial. Their scheme was adopted as the standard for control surfaces, but with modifications. Hinged rudders and elevators were fitted to fixed stabilizers, thus adding static stability while preserving control. Wing warping was replaced with simpler hinged ailerons. As aircraft grew from being purely sporting vehicles to a practical form of transport, designers moved from the Wright practice of negative static stability to positively stable designs that were able to fly steadily without constant control inputs.

By 1910 airplanes were typically capable of carrying one or two people at speeds of about 125 km/h for one hour, with applications chiefly for sport and military reconnaissance and observation. Their value in World War I led to consider-

able pressure for improvement. Engine outputs as great as 200 kilowatts allowed larger and more robust structures carrying loads of 400 kilograms or more at speeds of up to 200 km/h for several hours. Frameworks of welded steel tubing began to appear, and a few aircraft were sheathed in aluminum.

Before World War I, a novel type of airplane engine, the rotary, was developed. The cylinders were disposed radially, like spokes of a wheel, and the crankcase revolved with the propeller about a crankshaft fixed to the airframe. Both lubrication (using a once-through total-loss system) and cooling were thus improved, and rotaries were relatively lightweight (less than 2 kilograms per kilowatt) and reliable. The inertial forces of the whirling engine hindered application of more powerful rotaries, but this led to interest in the fixed radial configuration, with a stationary crankcase and rotating crankshaft. Once vibration problems were ameliorated, the radial became one of the two major airplane engine types later in the 1920s. The other was the water-cooled inline engine, often with two banks in a V8 or V12 configuration. By the end of the 1920s, outputs as great as 425 kW with weights as low as 1 kg/kW were becoming available.

Increasing engine power and a clearer appreciation of the demands of flight led to general advances in performance and to the beginnings of commercial air service in the 1920s. Externally braced monoplane wings became common, as did two or three engines for larger models. However, flight remained hazardous and limited to distances of less than 2,000 kilometers and speeds of less than 260 km/h, racers and stunts aside.

Although physicists had built an impressive body of theory about fluid flows in the eighteenth and nineteenth centuries, little of this was useful to early airplane designers. More relevant discoveries were quick in coming, particularly in Germany, under the leadership of Ludwig Prandtl. However, this work did not become generally known and accepted among designers until after World War I when it led to the adoption of thicker airfoil sections that provided better performance while allowing for stronger structures needing less drag-inducing bracing (Figure 1). However, overall flight efficiency, as measured by the ratio of lift to drag (L:D), improved little.

Achievements of the 1920s did not represent an ultimate limit on externally braced airplanes with fabric-covered tube structures, but development of the streamlined monoplane gradually led to the virtual extinction of the earlier aircraft. By the end of the twentieth century such construction was limited to light sport and special purpose aircraft.

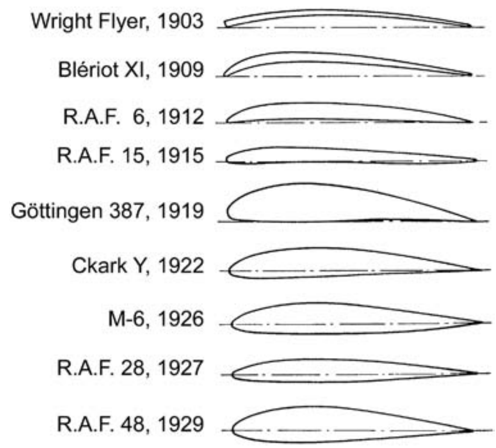


Figure 1. Evolution of airfoil sections, showing the trend to thicker airfoils as aerodynamic knowledge grew.

Streamlined Monoplanes

Before 1930, airplane development was little affected by science, and few designers knew much about the achievements of Ludwig Prandtl and his followers in understanding the physical mechanisms of lift and drag. Physics-based theories of structural strength, imported from architectural practice, found some use, but they were not readily applied to more elaborate aeronautical structures. In any event, the lack of knowledge of aerodynamic loads limited their use. The formula of the fabric-covered steel-tube structure with externally braced biplane or monoplane wings had become a comfortable one for many designers, and they saw the needs in terms of incremental improvement, not radical change.

As aerodynamicists drew on theory and experimental data to show a great gap between actual and potential performance, more and more designers became determined to close it. Particularly thought-provoking were findings that small struts and even bracing wires produced far more drag than had been supposed. Systematic scientific experimentation with wind tunnels, guided by the developing aerodynamic theory, stimulated many improvements in the late 1920s, of which cowlings that dramatically reduced the drag of radial engines while providing good cooling were among the most significant. Flaps and slats were devised to allow wing surfaces to be made smaller while still providing sufficient lift for take-off and landing at reasonable speeds.

Mechanisms were developed to permit undercarriages to be retracted.

Full implementation of aerodynamic improvements depended on structural advances. With new theories and extensive experimentation, designers learned to build airplane structures as smooth, integrally stiffened shells out of easily worked sheet aluminum fastened with rivets. They gained the knowledge and confidence to make wings of higher aspect ratio (the ratio of span to average chord length, from leading edge to trailing), without external bracing, thus improving lift efficiency. As speeds rose to 300 km/h and beyond, small details began to become quite significant. Engineers developed techniques of flush riveting in order to make even smoother structures.

These developments tended to make aircraft heavier, but weight was offset with power. By 1933, 800 kW was available with a weight about 0.8 kg/kW. By 1940, 1500 kW was available for 0.7 kg/kW. Radial and inline engines developed in parallel, with radials generally preferred for transports, bombers, and all naval aircraft, and inlines used for land-based fighters (with some exceptions in both directions). Pressurized cooling systems allowed reductions in radiator size and drag. Engine supercharging became widely used, either in mechanically driven or exhaust-turbine form, permitting aircraft to operate at higher altitudes where decreased atmospheric density reduced drag. In the U.S., development of high octane fuels permitted increased compression ratios.

The overall result was a remarkable transformation in the shape and performance of aircraft in less than a decade. Although much of the basic knowledge behind the transformation came from Germany, it was America that took the lead, reflecting favorable economic and geographic conditions and extensive government investment in practical research. British industry initially was somewhat slow to take up the new technologies, but it was quick to close the gap as rearmament began later in the 1930s.

The externally braced Ford AT-5 Trimotor transport of 1930, weighing 6100 kg and carrying 15 passengers at less than 215 km/h, had yielded by 1936 to the Douglas DC-3 of 11,400 kg, carrying 21 passengers in much greater comfort and capable of 370 km/h (although normal cruising speed was about 275 km/h). In Britain, fighters ranged from the 1930 Bristol Bulldog, with speeds reaching 280 km/h, to the 415 km/h Gloster Gladiator biplane that entered service as late as 1937, to the 585 km/h Supermarine Spitfire of 1939. If World War II had started in 1929, the planes that would have fought in it would have been only marginally superior to those of World War I. But by the time the war actually began in 1939, aircraft had moved far ahead and correspondingly played a much larger role.

Even more than in World War I, the pressures of conflict prompted great advances. By 1948, piston engine outputs had reached as high as 2600 kW, and some units could achieve 0.5 kg/kW. The North American P-51H fighter of 1945 could reach 780 km/h, while the 25,000 kg Lockheed Constellation airliner of 1946 carried 51 passengers at speeds up to 530 km/h, cruising at 440 km/h. The Constellation incorporated cabin pressurization, allowing comfortable operation at altitudes of 6000 m and more. Most of the performance gains of this period simply reflected increased engine power with incremental aerodynamic refinement.

The fundamental engineering science that underlay these achievements had all been in place by the mid-1930s. As early as the 1840s, physicists had formulated the exact equations for fluid flow (the Navier–Stokes equations). They were far too complex to be solved for any realistic cases, and even workable approximate solutions were very difficult to obtain. Efforts were made to analyze flow over airfoils with the assumption of incompressible, irrotational, and inviscid (friction-free) flow, but the results agreed poorly with measurements, especially for drag.

Prandtl and those who followed him showed that viscosity was in fact of crucial importance but that its effects were largely confined to a very thin layer of air, the "boundary layer," which normally lay immediately adjacent to the airplane's surface. Understanding of boundary layer phenomena was complicated by the existence of two quite different kinds of flow: smooth, relatively simple "laminar" flow and highly complex turbulent flow. Full understanding and prediction of turbulent flows remained elusive even at the end of the nineteenth century. Nevertheless, by making allowance for the boundary layer in their mathematics, aerodynamicists were able to reach approximate solutions to flow problems that were useful in practical cases, guiding designers to aerodynamic forms which could be refined with reasonable time and effort through experiments in wind tunnels.

The main problem was to gain lift by favorable distributions of air pressure and to avoid drag caused by adverse pressures. Pressure is constant for air at rest, but it is modified by the flow of air past the airplane in ways predicted by the approx-

imate solutions of the Navier–Stokes equations. Lift is generated by excess pressure on the underside of the airfoil relative to that on top. Some drag comes from viscous friction, but much is a result of the excess pressure on the front of airfoils and other parts of the plane over that acting on the rear. Particularly to be avoided are situations in which the flow separates entirely from the surface, leaving a relative void filled by irregular eddies; the low pressures thus generated result in high drag. Intensive application of this knowledge, involving large-scale testing and analysis in specialized facilities, directly supported aircraft and engine designers.

By 1950, the streamlined metal propeller-driven monoplane with piston engine power had nearly reached its inherent technological limits. Its triumph had been magnificent but brief. Gas-turbine power plants had now appeared, bringing new opportunities and new challenges. The classic subsonic airplane, with piston or turboprop engines, continued to have important market niches at the end of the twentieth century but only in subsidiary roles.

Transonic Aircraft

The aerodynamic theory that had been so valuable in the 1930s depended on the important simplifying assumption that compression of the air by the aircraft could be ignored. This assumption of incompressible flow broke down by 1940, as speeds approached Mach 0.7. (The Mach number is the ratio of the speed of the flow of air to that of sound. All moderate pressure disturbances travel at Mach 1 = 1225 km/h = 340 meters per second under standard conditions at sea level.)

The aircraft itself need not approach Mach 1 very closely to encounter compressibility, since the air accelerates in flowing past the airplane. Any advance beyond the flight conditions achieved by the end of World War II in 1945 could come only by penetrating this new "transonic" region of flight.

Before this became possible, new means of propulsion were necessary. The speed of a propeller blade through the air is, of course, higher than that of the aircraft itself. Therefore propellers were the first thing to be affected as sonic speeds were approached, with consequent falloff of performance. As the 1930s advanced, engineers in several countries turned to a new application of an old idea: a gas turbine to generate a hot gas stream for propulsion by jet reaction using atmospheric air, rather than carrying oxidizer on board the aircraft as in a rocket. These ideas were pursued in Britain and Nazi Germany, and late in World War II they were put into practice by both countries.

The stage was set for transonic flight if its problems could be overcome. With German aeronautical research and development prostrate following World War II, the lead passed principally to the U.S., Britain, and the Soviet Union, with the U.S. taking the most prominent role owing to its great resources. Quite rapid progress was made both in theoretical understanding of transonic flight (building to a significant extent on foundations laid down in Germany) and in the difficult problems of measurement. There were many important discoveries, but three deserve special mention. First was the use of wing sweepback. It is principally the component of airflow at right angles to the wing's leading edge that determines lift and drag, so sweep increases the wing's critical Mach number (the point at which its drag begins to rise very sharply due to compressibility). Second was the use of very thin sections with quite sharp leading edges, generally with little sweepback and low aspect ratio. Finally, area rules governed the overall design of transonic aircraft; that is, the aircraft is designed to ensure that the total cross section of all areas varies in as smooth a way as possible along its length, with no bumps or hollows, thus minimizing wave drag caused by formation of shock fronts. Through these and other means it was possible to raise the lift-to-drag ratios of large aircraft to 20:1 or more, offsetting the high fuel consumption of early jet engines and ultimately giving transonic aircraft unprecedented range–payload performance as engine efficiencies were improved.

High flight speeds increased the loads on the structures of transonic aircraft, as these vary according to the square of speed. The aerodynamic innovations posed significant additional challenges as they dictated greater deviations from structurally efficient forms. The problems were most notable in the wings, where depth had to be sacrificed and where sweepback increased effective structural aspect ratio. Designers met these challenges with structures employing complex internal geometries to optimize stiffness and strength. Highly stressed parts were frequently made of stainless steel or titanium, often in the form of massive forgings machined to net shape. These innovations raised the cost of aircraft construction significantly, but the advantages of transonic flight justified the expense for many applications.

The theoretical and experimental understanding of transonic aircraft aerodynamics grew to meet

the need for loads data on which to base structural design. It became clear that practical aircraft structures could not be even approximately rigid under transonic flight and maneuver loads, fostering much greater development of the discipline of aeroelasticity—analysis of the deformation of the plane's structure under aerodynamic loads and of the effect on aerodynamics (and hence on loads) of the change in its shape.

With the ratio of flight speed to stalling speed now reaching 4:1 or more, control problems took on new complexity. Controls that were effective and necessary at low speeds could become entirely ineffective or even counter-effective at transonic speeds. The classic example is the aileron on a long, swept wing, which had to be supplemented or supplanted with spoilers or other controls. Manual control circuits were replaced with powered irreversible controls.

In several cases, problems—particularly those associated with control and aeroelastic issues—were first recognized through catastrophic accidents, some of them in airliners in service. In earlier days it had been assumed that aviation was quite dangerous. However, as aircraft became increasingly employed for transport on a large scale, tolerance for accidents declined sharply. By the 1960s, even the military services had come to reject the human and economic costs of high loss rates. This emphasis on safety as well as reliability combined with the rigors of transonic flight to change the design of the airplane from something accomplished by ten to 25 engineers over a period of a year into a massive engineering project involving hundreds and even thousands of specialists over a number of years.

Transonic aircraft reached relative maturity by the 1960s, and most of the development in the last third of the century was in the nature of incremental improvements. Three significant developments do deserve mention however. First was the replacement of direct manual control with computer-mediated electronic "fly-by-wire" controls. Second was the introduction of fiber-reinforced materials, especially those incorporating carbon fibers in a thermosetting plastic matrix. After a lengthy gestation, occasioned both by concern for proven safety and the complexities of designing for a quite different kind of material, carbon-fiber composites finally began to see service in quantity in the final decade of the century, bringing significant benefits of weight reduction and promises of possible future reductions in fabrication costs (balanced against higher material costs). The third significant development was increased computerization of the design process, ranging from paperless drawings to complex calculations of aerodynamic flows. This has not, as once hoped, gone far in cutting the time or cost of design, but it has permitted unprecedented thoroughness, resulting in a better and ultimately more economical aircraft.

At the end of the twentieth century, the jet-propelled transonic aircraft was the dominant type for the great majority of transport services and for a wide range of military applications. There was no immediate prospect that it might be supplanted by supersonic or hypersonic aircraft in most roles, suggesting that transonic aircraft would continue to see wide use and evolve into the twenty-first century.

Supersonic and Hypersonic Aircraft

As the airflow over an aircraft reaches and exceeds Mach 1, drag begins to rise very steeply. Most transonic aircraft lack propulsion power to push very far into the drag rise in level flight, but by 1950 it had been verified that transonic aircraft could exceed flight speeds of Mach 1 in a dive.

Fighter aircraft need excess power for maneuver, generally achieved by adding afterburning to the jet engine. The next step, begun early in the 1950s, was to refine the aerodynamics and controls of fighters to permit them to fly at supersonic (over Mach 1) speeds in afterburning. The most notable changes involved adoption of greater sweep angle, highly swept delta wings, or thin, low-aspect wings with sharp leading edges. By 1960, most fighters entering service were capable of exceeding Mach 2. This was possible only at high altitudes (limited by aerodynamic heating and forces) and for brief periods (limited by high afterburning fuel consumption), but the speed was tactically useful. These were transonic airplanes that were capable of supersonic sprints.

As transonic aircraft entered service in substantial numbers for military and commercial purposes in the 1950s, it was generally anticipated that they would soon be supplanted or at least widely supplemented by truly supersonic aircraft that normally flew at over Mach 1. However, by 2000, only one type of aircraft regularly spent more than half of its time aloft in supersonic flight, the Anglo–French Concorde airliner, and only about a dozen Concordes remained in service.

Obstacles to wider supersonic flight included weight, cost, and environmental impact. Theory and experiment quickly led to the conclusion that the best shape for supersonic flight was slender and arrow-like and that suitable slender aircraft could cruise supersonically with efficiency generally

matching that of transonic aircraft. Slender airplanes were not inherently suited to the relatively low speeds needed for landing and take-off, however. Compromises and adaptations were necessary for controllable and efficient flight over a range of speeds that varied by 10:1 or more from maximum to stalling, leading to extra weight and expense. Moreover, supersonic flight presented even greater structural challenges than transonic flight, and this also brought cost and weight penalties. These arose in part from the high dynamic pressures involved in flight at very high speeds, but even more so from aerodynamic heating, representing the sum both of friction and of air compression in the supersonic flow. For sustained flight at more than Mach 2.5, aluminum loses too much strength due to heating to be used as a structural material unless it is cooled or protected. Steel or titanium may be used instead.

In aircraft, any increase in weight brings cascading problems. This is especially true for supersonic aircraft, which tend to be most attractive for long-range missions requiring large fuel loads. High weight allied with the need for special materials and structures pushed costs up for supersonic aircraft. Moreover the supersonic shock wave reaches disturbing and even destructive levels on the ground below the path of the supersonic plane even when it flies at altitudes of 20 kilometers or more. These problems combined to drastically slow acceptance of supersonic flight. Indeed, one supersonic type that did see successful service, the U.S. Lockheed SR-71 Mach 3+ strategic reconnaissance aircraft, was ultimately withdrawn from operations because its functions could be performed more economically by other means.

At over Mach 4, a series of changes in aerodynamic phenomena led to the application of the label "hypersonic." In principle, hypersonic flight presents attractive opportunities. In the 1960s there was a belief that supersonic aircraft might be supplemented relatively rapidly by hypersonic types. In practice, the problems of weight, cost, and environmental effects proved to be even more intractable. At hypersonic speed, heating is so intense that even steel and titanium lose strength. A number of research programs relating to hypersonic flight, stimulated in part by the demands of reentry from space, led to the accumulation of considerable knowledge of many of the issues. Progress on development of air-breathing propulsion systems was halting however, and several efforts aimed at construction of a prototype hypersonic aircraft collapsed owing to cost and technology issues. Thus at the end of the twentieth century, the promise of supersonic flight seemed just out of reach and that of hypersonic flight not yet clearly in view.

See also **Civil Aircraft, Jet-Driven; Civil Aircraft, Propeller-Driven; Civil Aircraft, Supersonic; Rocket Planes; Wright Flyers**

William D. O'Neil

Further Reading

Anderson, J.D. Jr. *A History of Aerodynamics and its Impact on Flying Machines*. Cambridge University Press, Cambridge, 1998.

Boot, R. *From Spitfire to Eurofighter: 45 Years of Combat Aircraft Design*. Airlife Publishing, Shrewsbury, 1990. A personal story of important designer of 1950s through 1970s illustrating design process and setting.

Boyne, W.J. *Technical Progress in World War I Aviation*. AIAA, Washington, 1978, 78–3002.

Boyne, W.J. and Donald S. Lopez, Eds. *The Jet Age: Forty Years of Jet Aviation*. National Air & Space Museum, Washington, 1979.

Brooks, P.W. *The Modern Airliner: Its Origins and Development*. Putnam, London, 1961.

Covert, E.E. Introduction, in *Thrust and Drag: Its Prediction and Verification,* Covert E.E., Ed. AIAA, Washington, 1985. Extended historical survey.

Dale, H. *Early Flying Machines*. British Library, London, 1992.

Fielding, J.P. *Introduction to Aircraft Design*. Cambridge University Press, Cambridge, 1999. Accessible modern survey requiring minimal technical preparation.

Foss, R.L and Blay, R. From Propellers to Jets in Fighter Design. *Lockheed Horizons,* 23, 2–24, 1987.

Gunston, B. *The Development of Pisto AeroEngines*. Patrick Stephens, Sparkford, 1993.

Howard, H.B. Aircraft Structures. *J. Aeronaut. Soc.,* 70, 54–66, 1966. Summary of development by a pioneer structural engineer.

Jarrett, P., Ed. *Aircraft of the Second World War: The Development of the Warplane, 1939–45*. Putnam, London, 1977.

Jarrett, P.P. *Biplane to Monoplane: Aircraft Development, 1919–39*. Putnam, London, 1997.

Kermode, A.C. *The Aeroplane Structure*. Pitman, London, 1964. Nontechnical survey.

Küchemann, D. *The Aerodynamic Design of Aircraft: A Detailed Introduction to the Current Aerodynamic Knowledge and Practical Guide to the Solution of Aircraft Design Problems*. Pergamon Press, Oxford, 1978. Authoritative survey looking ahead to supersonic and hypersonic aircraft.

Langley, M. *Metal Aircraft Construction: A Review for Aeronautical Engineers of the Modern International Practice in Metal Construction of Aircraft*. Pitman, London, 1937. Richly illustrated contemporary survey of metal aircraft development.

Loening, G.C. *Military Aeroplanes: An Explanatory Consideration of Their Characteristics, Performances, Construction, Maintenance, and Operation, for the Use of Aviators*. Best, Boston, 1917. Contemporary text summarizing state of art c. 1916. There are several editions.

Loftin, L.K. Jr. *Quest for Performance: The Evolution of Modern Aircraft*. NASA, Washington, (SP-486), 1985.

Miller, R. and Sawers D. *The Technical Development of Modern Aviation*. Praeger, New York, 1970.

Murphy, A.J. Materials in aircraft structures. *J. R. Aeronaut. Soc.,* 70, 114–119, Jan 1966. Historical survey by a pioneer.

Raymer, D.P. *Aircraft Design: A Conceptual Approach*, 3rd edn. AIAA, Washington, 1999. Comprehensive standard basic text with extensive bibliography.

Rhodes, T. History of the Airframe: Part 1. *Aerospace Engineering,* May 1989, 11–15.

Rhodes, T. History of the Airframe: Part 2. *Aerospace Engineering,* June 1989, 11–17.

Rhodes, T. History of the Airframe: Part 3. *Aerospace Engineering,* September 1989, 27–32.

Rhodes, T. History of the Airframe: Part 4. *Aerospace Engineering,* October 1989, 16–21.

Shenstone, B.S. Hindsight is always one hundred percent. *J. R. Aeronaut. Soc.,* 70 131–134, Jan 1966. Story of the streamlined monoplane revolution by an important designer of the period.

Shute, N. *Slide Rule: The Autobiography of an Engineer*. Heinemann, London, 1954. Personal story of important designer of the 1920s and 1930s.

Stinton, D. *The Anatomy of the Aeroplane*. Blackwell, Oxford, 1985. Another accessible modern survey with minimal technical demands.

Taylor, C.F. *Aircraft Propulsion: A Review of the Evolution of Aircraft Piston Engines*. National Air & Space Museum, Washington, 1971.

Warner, E.P. *Airplane Design: Aerodynamics*, 1st edn. McGraw-Hill, New York, 1927.

Warner, E.P. *Airplane Design: Performance*, 2nd edn. McGraw-Hill, New York, 1936. Two editions of a contemporary text illustrating the rapid progress between the late 1920s and 1930s.

Whitford, R. *Design for Air Combat*. Jane's, Coulsdon, Surrey, 1989. Deals extensively with aerodynamic issues of transonic and supersonic airplanes on a relatively nontechnical level; not restricted to fighters.

Whitford, R. *Fundamentals of Fighter Design*. U.K. Airlife Publishing, Shrewsbury, 2000. Surveys all aspects of fighter design from historical perspective on relatively nontechnical level; much of what is said applies to aircraft generally.

Aircraft Instrumentation

When Wilbur Wright took off from Kitty Hawk, North Carolina in December 1903, the only instruments on board the first successful powered heavier-than-air machine were a stopwatch and a revolution counter. The historical first flight began the quest to improve aircraft as machines to conquer the air, and instrumentation evolved accordingly.

Aircraft instrumentation can be divided into four categories: engine performance, communications, ground-based instruments, and air-based radar.

Engine Performance

Instruments to determine the performance of the plane were first to develop. The Wrights thought it was important to monitor engine revolutions to maintain engine performance. The stopwatch—which failed on the first flight—was intended to record the actual flight time for speed and endurance computations. Continued evolution of performance instrumentation record and report on the engine(s), on the airframe, and on the aircraft's attitude in the air. The first such instruments were timers and revolution counters (from the outset of heavier-than-air flight), airspeed indicators and compasses (from before World War I), as well as turn-and-bank indicators and artificial horizon instruments (during and after World War I). The six most important—and most prevalent—aircraft instruments are: altimeter (to gauge altitude), airspeed indicator (to gauge speed), the artificial horizon (to determine the attitude of the aircraft relative to the ground), the compass (for heading and direction finding), angle-of-attack indicator (to gauge the angle that the aircraft is climbing or diving), and the turn-and-bank indicator (to indicate the stability of the aircraft in a turn). Throughout their evolution, aircraft have included all of these in-plane instruments; the significant difference is that today most of these instruments are digital and computer enhanced rather than analog.

Communications Technology

Communications technology was the second system that developed in aircraft instrumentation. Initially too heavy for the primitive machines, wireless transmission sets (W/T or radio) were not installed in aircraft on a regular basis until well into World War I. The need was foreseen, but technological development had to catch up with requirements. When radios were installed, pilots could communicate with other planes and ground stations transmitting information back and forth. This had many important effects including military observation and coordination as well as increased safety in the air. Radio beams were also used introduced in the 1940s as directional beacons to pinpoint specific locations. They were used initially to guide planes to landing sites (the Lorenz system). Later, radio direction finding was adapted to guide bombers to distant targets. Both the Germans (*Knickebein* and *X-Gerät*) and the British (*Gee* and *Oboe*) used radio beams extensively to aid in targeting enemy cities during World War II. Radio direction finding is still used

extensively to guide planes to their destinations through instrument flight rules (IRF).

Ground-Based Instruments

The third interconnected evolution in aircraft instrumentation was the ability of ground-based instruments to locate and identify aircraft. In the 1930s the British and Germans developed ground-based radio-direction finding (radar) to be able to locate aircraft from a distance. The British also added the component of IFF (identification, friend or foe) beacons in aircraft to be able to identify British Royal Air Force planes from their German Luftwaffe attackers. The "Battle of Britain" in 1940 was won by the British through the use of Fighter Command: an integrated system of ground-based radar, landline communications, a central command system, and fighter squadrons. Ground-based radar has evolved to the point where air traffic controllers the world over maintain close contact with all aircraft in order to ensure safety in the air. Ground-based radar, although not specifically an aircraft instrument, is important for the operation of large number of aircraft in the skies today.

Air-Based Radar

The final example of important aircraft instrumentation also made its debut in wartime. Air-based radar—radar instruments in aircraft—was first used by the British RAF Bomber Command to locate German cities at night. The British were able to find and bomb cities using airborne radar. To counter Bomber Command, the German Luftwaffe devised *Lichtenstein*, airborne radar used to find RAF bombers in the air. To this day, modern military air forces use airborne radar to locate and identify other aircraft. However, not all uses are so nefarious. Airborne radar has been adapted to not only "see" other planes in the air, but also to find weather fronts in dense cloud so that bad weather can be avoided. Airborne radar has made commercial aviation safer than ever.

Targeting Instruments

In the military sphere, one additional evolution is in targeting instruments. From the early iron gunsights and dead-reckoning bombsights, instruments have been developed for military aircraft to deliver weapons payloads with ever-increasing efficiency. During World War I, pilots had to rely on skill alone to deliver bombs and bullets against enemy targets. The first computerized gunsights were developed during World War II, as well as the American Norden bombsight—the bombsight developed for the American "precision" bombing campaign against Germany. Heads-up-displays (or HUDs) in the cockpit have increased the accuracy of weapons delivery into the modern era with the addition of laser-guided munitions and "smart" bombs. The modern HUD and complementary bombsights can track a number of targets, which adds to the destructive efficiency of modern-day military aircraft.

During World War II Albert Rose, Paul K. Weimer and Harold B. Law at RCA developed a small experimental compact image orthicon camera system to guide GB-4 glide bombs to their target (see Figure 2). A television camera carried under the bomb's fuselage transmitted a radio signal that was picked up by an antenna on the controlling plane, before being displayed as an image on a cathode ray tube. The image was displayed to the bombardier, who could then send radio commands to correct the glide bomb's course.

Additional Instrumentation

Additional instrumentation has been added to aircraft at different times to record high altitude,

Figure 2. Guided missile cam, WWII.
[*Courtesy of the David Sarnoff Library*.]

high-speed flight, atmospheric conditions, and global navigation. In the modern era, pilots can rely on global positioning satellites (GPS) to maintain precise flight paths. However, even to this day, pilots of both fixed-wing and rotary-wing aircraft rely on simple instruments—sometimes disguised as complicated computer enhanced systems—such as the aircraft compass, engine(s) gauges, artificial horizon, turn-and-bank indicator, flight clock, radio and direction finding aids. The most important instruments that pilots possess, however, may still be instinct, common sense, and seat-of-the-pants daring.

See also **Fly-by-Wire Systems; Global Positioning System (GPS); Gyro-Compass and Inertial Guidance; Radar Aboard Aircraft; Radionavigation**

MIKE PAVELEC

Further Reading

Beij, K.H. *Air Speed Instruments*. Government Printing Office, Washington, D.C., 1932.

Dwiggins, D. *The Complete Book of Cockpits*. Tab Books, Blue Ridge Summit, PA, 1982.

Gracey, W. *Measurement of Aircraft Speed and Altitude*. Wiley, New York, 1981.

Hersey, M. *Aeronautic Instruments, General Classification of Instruments and Problems*. Government Printing Office, Washington, D.C., 1922.

Hersey, M. *Aeronautic Instruments, General Principles of Construction, Testing, and Use*. MIT Press, Cambridge, MA, 1922.

Hunt, F. *Aeronautic Instruments*. Government Printing Office, Washington, D.C., 1923.

Irvin, G. *Aircraft Instruments*. McGraw-Hill, New York, 1944.

Molloy, E. *Aeroplane Instruments*. Chemical Publishing, New York, 1940.

Nijboer, D. *Cockpit*. Boston Mills Press, Toronto, 1998.

Pallett, E.H.J. *Aircraft Instruments, Principles and Applications*. Pitman, London, 1981.

Paterson, J. *Aircraft Compass Characteristics*. National Advisory Committee for Aeronautics (NACA), Washington, D.C., 1936.

Aircraft Instruments, United States Naval Air Technical Training Command, Washington, D.C., 1954.

Alloys, Light and Ferrous

Although alloys have been part of industrial practice since antiquity (bronze, an alloy of copper and tin, has been in use for thousands of years), the systematic development of alloys dates to the middle of the nineteenth century. Due to improvements in techniques for chemical analysis and the rise of systematic testing of material properties, the basic theoretical principles of alloys were developed in the late 1800s.

Broadly speaking, the development and use of alloys in the twentieth century was essentially an extension of the discoveries made in the nineteenth century. Refinements in the purity of source materials, the systematic testing of different alloy combinations and methods of heat treatment, and improvements in manufacturing techniques led to significant improvements in material properties. However, no new fundamental principles that led to radical breakthroughs were discovered during the twentieth century.

During the twentieth century, steel was the dominant engineering material in the industrialized world due to its low cost and versatility. Much of that versatility is due to a class of steels known as alloy steels. By adding other metals to the basic mix of iron and carbon found in steel, the properties of alloy steels can be varied over a wide range. Alloy steels offer the potential of increased strength, hardness, and corrosion resistance as compared to plain carbon steel. The main limitation on the use of alloy steels was that they typically cost more than plain carbon steels, though that price differential declined over the course of the twentieth century.

Although steel was the dominant engineering material in the twentieth century, a number of other alloys developed during the twentieth century found widespread use in particular applications. Higher cost limited the use of specialty alloys to particular applications where their material properties were essential for engineering reasons.This entry covers the use of alloys in mechanical applications. Alloys used in electrical applications are discussed in the entry Alloys, Magnetic.

Definitions

An alloy is a mixture of two or more metallic elements or metallic and nonmetallic elements fused together or dissolving into one another when molten. The mixture is physical, and does not involve the formation of molecular bonds. Strictly speaking, steel is an alloy, since it is a mixture of iron and carbon, but is not normally referred to in that way. Rather, when one speaks of alloy steel, one is referring to steel (iron plus carbon) with other elements added to it.

The formal definition of alloy steel is a steel where the maximum range of alloying elements content exceeds one or more of the following limits: 1.6 percent manganese, 0.6 percent silicon, or 0.6 percent copper. In addition, alloy steels are recognized as containing specific (minimum or otherwise) quantities of aluminum, boron, chro-

mium (up to 3.99 percent), cobalt, nickel, titanium, tungsten, vanadium, zirconium, or any other alloying element that is added in order to obtain a desired alloying effect.

Somewhat confusingly, a number of alloys that are commonly referred to as alloy steels actually contain no carbon at all. For example, maraging steel is a carbon-free alloy of iron and nickel, additionally alloyed with cobalt, molybdenum, titanium and some other elements.

Another commonly used industry term is "special" (or in the U.S. "specialty") steel. Most, though not all, special steels are alloy steels, and the two terms are often used interchangeably. Other industry terms refer to the properties of the steel rather than a specific material composition. For example, "high strength" steel refers to any steel that can withstand loads of over 1241 MPa, while "tool-and-die" steel refers to any steel hard enough to be used for cutting tools, stamping dies, or similar applications.

The names of nonsteel alloys are usually defined by the names of their primary constituent metals. For example, nickel–chromium alloy consists of a mix of approximately 80 percent nickel and 20 percent chromium. Titanium alloys are primarily titanium mixed with aluminum, vanadium, molybdenum, manganese, iron or chromium. However, some alloys are referred to by trade names that have become part of the standard engineering vocabulary. A good example is Invar, an alloy of 64 percent iron and 36 percent nickel. The name is a contraction of the word "invariable," reflecting Invar's very low rate of thermal expansion.

Alloys are useful for industrial purposes because they often possess properties that pure metals do not. For example, titanium alloys have yield strengths up to five times as high as pure titanium, yet are still very low in density. Even when alloys have the same properties as pure materials, alloy materials—particularly alloy steels—are often cheaper than a pure material for a given purpose.

The differences in properties between a pure material and its alloys are due to changes in atomic microstructure brought about by the mixture of two or more types of atoms. The addition of even small amounts of an alloying element can have a major impact on the arrangement of atoms in a material and their degree of orderly arrangement. In particular, alloying elements affect the way dislocations are formed within microstructures. These changes in microstructure lead to large-scale changes in the properties of the material, and often change the way a material responds to heat treatment.

It is important to note that the addition of alloying elements can have both positive and negative effects on the properties of a material from an engineering point of view. In the manufacture of alloys, it is often just as important to avoid or remove certain chemical elements as it is to add them. Careful control of the chemical composition of raw materials and various processing techniques are used to minimize the presence of undesirable elements.

Alloy Steel

The development of alloy steel has its origins in the crucible process, perfected by Benjamin Huntsman in England around 1740. By melting bar iron and carbon in clay pots and then pouring ingots, Huntsman created superior steel with carbon uniformly dispersed throughout the metal. Used for cutlery, die stamps and metal-cutting tools; crucible steel was the first specialty steel.

In 1868, Robert F. Mushet, the son of a Scottish ironmaster, found that the addition of finely powdered tungsten to crucible steel while it was melted made for much harder steel. Suitable for metal cutting tools that could operate at high speed, Mushet tungsten tool steel was the first commercial alloy steel. The English metallurgist and steelmaker Sir Robert Hadfeld is generally considered to be the founder of modern alloy steel practice, with his invention of manganese steel in 1882. This steel, containing 12 percent manganese, has the property of becoming harder as it is worked. This made it ideal for certain types of machinery, such as digging equipment. Hadfeld also invented silicon steel, which has electrical properties that make it useful for building transformers. His work showed conclusively that the controlled addition of alloying elements to steel could lead to significant new specialty products. Hadfeld's discoveries, which were well publicized, led many other engineers and steelmakers to experiment with the use of alloying elements, and the period between about 1890 and 1930 was a very active one for the development of new alloys.

The first highly systematic investigation of alloy steels was carried out by Frederick W. Taylor and Maunsel White at the Bethlehem Steel Works in the 1890s. In addition to testing various alloy compositions, the two men also compared the impact of different types of heat treatment. The experiments they conducted led to the development of high-speed steel, an alloy steel where tungsten and chromium are the major alloying elements, along with molybdenum, vanadium and cobalt in

varying amounts. These steels allowed the development of metal cutting tools that could operate at speeds three times faster than previous tools. The primary application of high-speed steel during the twentieth century was for the manufacture of drill bits.

Military applications were also a major factor in the development of alloy steels. The demand for better armor plate, stronger gun barrels, and harder shells capable of penetrating armor led to the establishment of research laboratories at many leading steel firms. This played a significant role in the development of the science of metallurgy, with major firms like Vickers in the U.K. and Krupps in Germany funding metallurgical research. The most notable discovery that came out of this work was the use of nickel as an alloying element. Nickel in quantities between 0.5 and 5.0 percent increases the toughness of steel, especially when alloyed with chromium and molybdenum. Nickel also slows the hardening process and so allows larger sections to be heat-treated successfully.

The young science of metallurgy gradually began to play a greater role in nonmilitary fields, most notably in automotive engineering. Vanadium steel, independently discovered by the metallurgists Kent Smith and John Oliver Arnold of the U.K. and Léon Guillet of France just after the beginning of the twentieth century, allowed the construction of lighter car frames. Research showed that the addition of as little as 0.2 percent vanadium considerably increased the steel's resistance to dynamic stress, crucial for car components subject to the shocks caused by bad roads. By 1905, British and French automobile manufacturers were using vanadium steel in their products. More significantly, Henry Ford learned of the properties of vanadium from Kent Smith and used vanadium alloy steel in the construction of the Model T. Vanadium steel was cheaper than other steels with equivalent properties, and could be easily heat-treated and machined. As a result, roughly 50 percent of all the steel used in the original Model T was vanadium alloy. As the price of vanadium increased after World War I, Ford and other automobile manufacturers replaced it with other alloys, but vanadium had established the precedent of using alloy steel. By 1923, for example, the automobile industry consumed over 90 percent of the alloy steel output of the U.S., and the average passenger car used some 320 kilograms of alloy steel.

The extensive use of alloy steels by the automobile industry led to the establishment of standards for steel composition. First developed by the Society of Automotive Engineers (SAE) in 1911 and refined over the following decade, these standards for the description of steel were widely adopted and used industry-wide by the 1920s, and continued to be used for the rest of the century. The system imposes a numerical code, where the initial numbers described the alloy composition of the steel and the final numbers the percentage of carbon in the steel. The specifications also described the physical properties that could be expected from the steel, and so made the specification and use of alloys steels much easier for steel consumers.

One of the goals of automotive engineers in the 1910s and 1920s was the development of so-called "universal" alloy steel, by which they meant a steel that would have broad applications for engineering purposes. While no one alloy steel could serve all needs, the search for a universal steel led to the widespread adoption of steel alloyed with chromium and molybdenum, or "chrome–moly" steel. This alloy combines high strength, toughness, and is relatively easy to machine and stamp, making it the default choice for many applications.

The final major class of alloy steel to be discovered was stainless steel. The invention of stainless steel is claimed for some ten different candidates in both Europe and the U.S. in the years around 1910. These various individuals all found that high levels of chromium (12 percent or more) gave exceptional levels of corrosion resistance. The terms "stainless" is a bit of an exaggeration—stainless steel alloys will corrode under extreme conditions, though at a far slower rate than other steels. It is this resistance to corrosion, combined with strength and toughness, that made stainless steels so commercially important in the twentieth century. The first commercial stainless steels were being sold by 1914 for use in cutlery and turbine blades, and by the 1920s the material was commonly used in the chemical industry for reactor vessels and piping. Stainless steel later found widespread application in the food processing industry, particularly in dairy processing and beer making. By the end of the twentieth century, stainless steel was the most widely produced alloy steel.

After the 1920s, the development of alloys steels was largely a matter of refinement rather than of significant new discoveries. Systematic experimentation led to changes in the mix of various alloys and the substitution of one alloy for another over time. The most significant factor has been the cost and availability of alloying elements, some of which are available in limited quantities from

only a few locations. For example, wartime shortages of particular elements put pressure on researchers to develop alternatives. During World War II, metallurgists found that the addition of very small amounts of boron (as little as 0.0005 percent) allowed the reduction of other alloying elements by as much as half in a variety of low- and medium-carbon steels. This started a trend that continued after the war of attempts to minimize the use of alloying elements for cost reasons and to more exactly regulate heat treatment to produce more consistent results.

The manufacture of alloy steels changed significantly over the period 1900–1925. The widespread introduction of electrical steel making replaced the use of crucible furnaces for alloy steel processing. Electrical furnaces increased the scale of alloy steel manufacture, and allowed the easy addition of alloying elements during the steel melt. As a result, steel produced electrically had a uniform composition and could be easily tailored to specific requirements. In particular, electric steel-making made the mass production of stainless steel possible, and the material became cheap enough in the interwar period that it could be used for large-scale applications like the production of railway cars and the cladding of the Chrysler and Empire State skyscrapers in New York.

A major refinement in steel manufacture, vacuum degassing, was introduced in the 1950s and became widespread by the 1970s. By subjecting molten steel to a strong vacuum, undesirable gases and volatile elements could be removed from the steel. This improved the quality of alloy steel, or alternatively allowed lower levels of alloy materials for the same physical properties.

As a result of manufacturing innovations, alloy steel gradually became cheaper and more widely used over the twentieth century. As early as the 1960s, the distinction between bulk and special steel became blurred, since bulk steels were being produced to more rigid standards and specialty steels were being produced in larger quantities. By the end of the twentieth century, nearly half of all steel production consisted of special steels.

Other Alloys

A variety of nonsteel alloy materials were developed during the twentieth century for particular engineering applications. The most commercially significant of these were nickel alloys and titanium alloys. Nickel alloys, particularly nickel–chromium alloys, are particularly useful in high temperature applications. Titanium alloys are light in weight and very strong, making them useful for aviation and space applications. The application of both materials was constrained largely by cost, and in the case of titanium, processing difficulties.

Nickel–chromium alloy was significant in the development of the gas turbine engine in the 1930s. This alloy—roughly 80 percent nickel and 20 percent chromium—resists oxidation, maintains strength at high temperatures, and resists fatigue, particularly from enbrittlement. It was later found that the addition of small amounts of aluminum and titanium added strength through precipitation hardening. The primary application of these alloys later in the twentieth century was in heating elements and exhaust components such as exhaust valves and diesel glow-plugs, as well as in turbine blades in gas turbines.

Pure titanium is about as strong as steel yet nearly 50 percent lighter. When alloyed, its strength is dramatically increased, making it particularly suitable for applications where weight is critical. Titanium was discovered by the Reverend William Gregor of Cornwall, U.K., in 1791. However, the pure elemental metal was not made until 1910 by New Zealand born American metallurgist Matthew A. Hunter. The metal remained a laboratory curiosity until 1946, when William Justin Kroll of Luxembourg showed that titanium could be produced commercially by reducing titanium tetrachloride (TiCl4) with magnesium. Titanium metal production through the end of the twentieth century was based on this method.

After the World War II, U.S. Air Force studies concluded that titanium-based alloys were of potentially great importance. The emerging need for higher strength-to-weight ratios in jet aircraft structures and engines could not be satisfied efficiently by either steel or aluminum. As a result, the American government subsidized the development of the titanium industry. Once military needs were satisfied, the ready availability of the metal gave rise to opportunities in other industries, most notably chemical processing, medicine, and power generation.

Titanium's strength-to-weight ratio and resistance to most forms of corrosion were the primary incentives for utilizing titanium in industry, replacing stainless steels, copper alloys, and other metals. The main alloy used in the aerospace industry was Titanium 6.4. It is composed of 90 percent titanium, 6 percent aluminum and 4 percent vanadium. Titanium 6.4 was developed in the 1950s and is known as aircraft-grade titanium. Aircraft-grade titanium has a tensile

strength of up to 1030 MPa and a Brinell hardness value of 330. But the low ductility of 6.4's made it difficult to draw into tubing, so a leaner alloy called 3-2.5 (3 percent aluminum, 2.5 percent vanadium, 94.5 percent titanium) was created, which could be processed by special tube-making machinery. As a result, virtually all the titanium tubing in aircraft and aerospace consists of 3-2.5 alloy. Its use spread in the 1970s to sports products such as golf shafts, and in the 1980s to wheelchairs, ski poles, pool cues, bicycle frames, and tennis rackets.

Titanium is expensive, but not because it is rare. In fact, it is the fourth most abundant structural metallic element in the earth's crust after aluminum, iron, and magnesium. High refining costs, high tooling costs, and the need to provide an oxygen-free atmosphere for heat-treating and annealing explain why titanium has historically been much more expensive than other structural metals.

As a result of these high costs, titanium has historically been used in applications were its low weight justified the extra expense. At the end of the twentieth century the aerospace industry continued to be the primary consumer of titanium alloys. For example, one Boeing 747 uses over 43,000 kg of titanium.

See also **Alloys, Magnetic; Iron and Steel Manufacture**

MARK CLARK

Further Reading

Freytag D.A. and Burton L.A., Eds. *The History, Making, and Modeling of Steel*. William K. Walthers, Milwaukee, 1996.

Misa, T.J. *A Nation of Steel: The Making of Modern America, 1865—1925*. Johns Hopkins University Press, Baltimore, 1995.

Seely B., Ed. *Iron and Steel in the Twentieth Century*. Facts on File, 1994

Smith, W.F. *Structure and Properties of Engineering Alloys*. McGraw-Hill, New York, 1993.

Williams, J.C. *et al. Titanium*. Springer-Verlag, 2003.

Alloys, Magnetic

The development and application of magnetic alloys in the twentieth century was driven largely by changes in the electronics industry. Magnetic alloys are used in three major applications. First, permanent magnets are required for devices like electric motors and generators, loudspeakers, and television tubes that use constant magnetic fields. Second, electrical steels are used to make electromagnets, solenoids, transformers and other devices where changing magnetic fields are involved. Finally, they are used in magnetic storage media, which require magnetic materials that retain the impression of external magnetic fields.

The most important types of magnetic materials are the ferromagnets, of which the most commonly used are iron, cobalt, and nickel. Ferromagnets have high magnetic ability, which allows high magnetic inductions to be created using magnetic fields. They also retain magnetism so they can be used as a source of field in electric motors or for recording information. Ferromagnets are used in the manufacture of electrical steels and magnetic media.

In the early twentieth century, the most commonly used magnetic materials were steel alloys containing tungsten or chromium. Chromium steel came to dominate the market due to lower cost. In 1917, Honda and Takai found that the addition of cobalt doubled the coercivity of chromium steel. Cobalt–chromium steels were commercialized in the early 1920s. The first major permanent magnet alloy, Alnico, was discovered in 1930 by Mishima and commercially introduced in the late 1930s. An alloy of steel, aluminum, nickel, and cobalt, Alnico has magnetic properties roughly eight times better than chromium–cobalt steels.

Introduced in the 1950s, ferrites—ceramic ferromagnetic materials—are a class of magnets made from a mixture of iron oxide with other oxides such as nickel or zinc. Ferrites have greatly increased resistivity because they are oxides rather than alloys of metals. They are also very hard, which is useful in applications where wear is a factor, such as magnetic recorder heads. Unlike bulk metals, ceramic magnets can be molded directly. Although not as strong on a unit weight basis as Alnico, ferrite magnets are much cheaper, and account for the vast majority of magnets used in industry in the late twentieth century—roughly 90 percent by weight in the 1990s, for example.

The strongest magnetic materials are the "rare-earth" magnets, produced using alloys containing the rare earth elements samarium and neodymium. Samarium–cobalt magnets were introduced in the early 1970s, but increases in the price of cobalt due to unrest in Zaire limited their use. In 1983, magnets based on a neodymium–iron–boron alloy were introduced. Neodymium is cheaper and more widely available than samarium, and neodymium–iron–boron magnets were the strongest magnetic materials available at the end of the twentieth century.

However, not all applications require the strongest magnetic field possible. Throughout the twentieth century, electromagnets and electromag-

netic relays were constructed almost exclusively from soft iron. This material responds rapidly to magnetic fields and is easily saturated. It also has low remnance (a measure of how strong a remaining magnetic field is), so there is little residual field when the external magnetic field is removed.

For transformers, the material properties desired are similar but not identical to those for electromagnets. The primary additional property desired is low conductivity, which limits eddy current losses. First developed just after 1900, the primary material used for power transformers is thus a silicon–iron alloy, with silicon accounting for approximately 3 to 4 percent by weight. The alloy is heat-treated and worked to orient the grain structure to increase permeability in a preferred direction. Aluminum–iron alloy is also a suitable material for this application, although it is less used due to its higher cost.

Transformers for applications that involve audio and higher frequencies make use of nickel–iron alloys, with a nickel content of 30 percent or more. Common trade names for such alloys include Permalloy, Mumetal, and Supermalloy. These alloys were developed in the early twentieth century and were first manufactured in quantity for use as submarine cable shielding. The decline in cable production in the 1930s led to their use in transformers and related applications.

Magnetic recording was first developed by the Danish inventor Valdemar Poulsen at the beginning of the twentieth century and used for sound recording. The first material used for magnetic recording, solid steel wire or tape, was originally developed for other applications. For example, steel piano wire was used as the recording media for early wire recorders. However, the property that makes particular steel alloys suitable for magnetic recording, strong remnant magnetism, is associated with increased stiffness. Thus, a recording tape or wire made from magnetic alloy steel is highly resistant to bending. This creates difficulties in the design of a mechanism for moving the recording media past the recording head.

As a result, by the late 1930s most recording media were divided into two parts. The first part was a suitable substrate, such as brass wire or plastic tape that could be fed easily through a reel or cassette mechanism. The second part was a coating that had suitable magnetic properties for recording. By the late 1940s, it was clear that the cheapest and most easily used recording media for sound recording was plastic tape coated with particles of a type of iron oxide (gamma ferric oxide). This type of tape continued in use through the end of the twentieth century due to its low cost. During the 1970s, new tape particles of chromium dioxide and cobalt-doped ferric oxide were introduced because of their superior magnetic properties, but their higher cost meant that they were used only for more specialized audio recording applications.

Magnetic media based on ferric oxide particles were used for the recording of computer data beginning in the 1950s. Initial computer recording applications used coated plastic tape. Metal disks coated with iron oxide were introduced in the late 1950s for use in computer disk drives. In the 1990s, thin metal films of cobalt alloy largely replaced metal oxide as the recording media for hard disks.

In addition to ferromagnetic materials, two additional classes of magnetic alloys exist: paramagnets and diamagnets. Aside from use in the scientific study of magnetism, paramagnets have limited uses. One limited application is the production of very low temperatures. Paramagnetic salts can be cooled conventionally and then demagnetized, producing temperatures in the millikelvin range.

Superconducting materials are a subclass of diamagnets. When cooled to a sufficiently low temperature, superconducting materials experience a significant drop in resistance. Associated with this transition is the exclusion of magnetic flux from the conductor, with the flux moving to the surface of the conductor. These properties allow for the production of very high magnetic fields when using niobium–tin alloys as the conducting material. These materials were also used in the development of magnetic resonance imaging (MRI) devices, although in the 1990s superconducting magnets were being replaced in this application by neodymium–iron–boron permanent magnet systems.

See also **Audio Recording, Compact Disk; Computer Memory; Materials and Industrial Processes**

MARK CLARK

Further Reading

Daniel, E.D., Mee, C.D. and Clark M.H. *Magnetic Recording: The First 100 Years*. IEEE Press, New York, 1999.

Jiles, D. *Introduction to Magnetism and Magnetic Materials*. Chapman & Hall, London, 1991.

Livingston J.D. *Driving Force: The Natural Force of Magnets*. Harvard University Press, Cambridge, MA, 1996.

Sowter G.A.V. Soft magnetic materials for audio transformers: history, production, and applications. *J. Aud. Eng. Soc.*, 35, 10, October, 1987.

Analgesics

Analgesics are drugs that relieve pain selectively, without affecting consciousness or sensory perception. This is the important distinction between analgesics and anesthetics. Intensive research on the biochemistry of analgesia from the 1970s has shown that there are two main classes of analgesics. The first, opioids, combine chemically with molecular "receptors" in the brain to block pain impulses in the central nervous system. Nonsteroidal anti-inflammatory drugs (NSAIDs) alleviate pain by inhibiting the production of prostaglandins, hormone-like substances that cause local inflammation and pain at the site of an injury or infection. Opioid analgesics can be used for short-term or long-term relief of severe pain; NSAIDs are used to relieve moderate pain, such as headaches, superficial injuries, or muscle strain.

Opium, the dried, powdered sap of unripe poppy seedpods, has been used as an analgesic since antiquity. Over 20 different alkaloids are found in dry opium, of which the most important is morphine. The correct chemical structure of morphine was proposed in 1925 and confirmed by total synthesis in 1955. Morphine, administered by subcutaneous injection, effectively relieves pain, but there are various side-effects such as drowsiness, respiratory depression, nausea, and vomiting. The analgesic effect peaks at about 1 hour after administration and lasts 4–5 hours. Morphine users develop physical dependence on the drug and become tolerant to it. Increasing doses are needed to maintain its effectiveness, and serious withdrawal symptoms accompany the cessation of morphine medication. Codeine, another naturally occurring opium alkaloid, is a less potent analgesic than morphine, but it is also less addictive and produces less nausea than morphine. It can be administered orally and is often used in conjunction with aspirin. Meperidine (Demerol), a synthetic morphine analog identified in 1939, was originally thought to provide short-lasting analgesia without addiction, but this proved false and the drug was widely abused. It is still the most common drug used for pain relief in childbirth and has superseded morphine. The side effects of meperidine are similar to those of morphine but are less severe and are further reduced when given with the antihistamine drug promethazine. In prolonged use meperidine may cross the placental barrier, and the drug has been found in newborn infants. Heroin, introduced in 1898, was also falsely heralded as a nonaddictive alternative to morphine. It is about ten times more potent than morphine when administered intravenously and is still used clinically for its analgesic properties in some countries, though it is not generally available for therapeutic purposes due to its highly addictive propensities. Methadone, a synthetic opioid analgesic discovered in Germany during World War II, has long-lasting analgesic effects when taken orally, and, as it alleviates the euphoria-producing effects of heroin, it is used to control the withdrawal symptoms of heroin addiction. The opioid analgesics were formally called narcotic drugs as they induce sleep and cause physiological dependence and addiction. The term is less commonly used in medicine because many drugs other than opioids also show these effects.

In 1973 complex compounds called opioid receptors, which can combine with opioid molecules, were discovered in the brain, hypothalamus, and spinal cord. At least eight such substances are known, though only four are thought to be important to the central nervous system. The best understood is the μ receptor, which affects euphoria, respiratory depression, tolerance, and analgesia. The κ receptor is also involved in analgesia as well as diuresis, sedation, and physical dependence. Soon after discovery of these receptors, peptide-like compounds consisting of chains of amino-acid residues and showing opioid properties were found in the pituitary gland. Three groups of endogenous opioid peptides known as endorphins, enkephalins, and dynorphins were discovered around 1975. Found to be neurotransmitters, one of the most important is β-endorphin, consisting of a chain of 30 amino acid residues. It is synthesized, stored, and released from cells in the pituitary gland, and it can also be prepared synthetically. Injected intravenously, β-endorphin is three times more potent than morphine. Interest in these drugs intensified when two potent analgesic pentapeptides, each containing five linked amino acids, were found in extracts of pig brain. Named enkephalins, they are derived from endorphins. They can be prepared synthetically and injected intravenously to induce analgesia by combining with receptors in the brain in the manner of the opioids.

Several chemically unrelated groups of organic compounds also show mild analgesic properties. They include derivatives of salicylic acid, pyrazolone, and phenacetin. These nonopioid drugs are often self-prescribed, though continued use can lead to adverse effects including drug abuse and addiction, allergic reactions, gastrointestinal irritation, and fatal overdose. The oldest and most

widely used nonopioid analgesic is acetyl salicylic acid, or aspirin, first developed and marketed in 1899 by the German chemical company, Bayer. In addition to its analgesic properties, aspirin reduces fever and inflammation by inhibiting the synthesis of prostaglandins. The irritant effect of large doses of aspirin on the stomach lining may result in gastric ulcers. Hypersensitivity to aspirin and related drugs is thought to be due to the accumulation of prostaglandins after the pathways that break them down are blocked. All aspirin-like analgesics inhibit prostaglandin synthesis, and their potency depends on the degree to which they do so. Many share similar side effects, some of which can be serious. However, the inhibition of prostaglandins also reduces the ability of blood platelets to form clots and this effect has given aspirin added value as an antithrombotic drug.

As the mechanisms of analgesic action began to be understood, alternatives to aspirin were introduced. Acetaminophen, a derivative of phenacetin introduced in 1956, is a popular alternative drug that avoids severe symptoms of stomach irritation. This mild analgesic and antipyretic however has much weaker anti-inflammatory properties, and overdoses can cause liver and kidney damage. Pyrazolone analgesics such as phenylbutazone show similar properties to aspirin and have been used to treat rheumatoid arthritis. Recently potent NSAIDs, sometimes called "super-aspirins," are widely used to replace aspirin itself. These include the propionic acid derivatives, naproxen (first made by Syntex in 1979), ketoprofen (by Wyeth in 1986) and ibuprofen. The latter, manufactured by Upjohn in 1984, is much more potent than aspirin, causes fewer side effects, and is better tolerated by most individuals. In large doses or prolonged use however, it can cause all the symptoms of other inhibitors of prostaglandin synthesis including gastric irritation and renal toxicity.

Further Reading

Atkinson, R.S., Rushman, G.B. and Lee J.A. *A Synopsis of Anaesthesia*, 10th edn. IOP Publications, Bristol, 1987. Covers all aspects of anesthesia and analgesia as practiced in the late twentieth century.

Bonica, J.J. *et al.,* Eds. *Advances in Pain Research and Therapy*. Raven, New York, 1983, 5, 199–208 and 411–459.

Fairley, P. *The Conquest of Pain*. Michael Joseph, London, 1978

Holden, A.V. and Winlow, W., Eds. *The Neurobiology of Pain*. Manchester University Press, Manchester, 1984.

Kirk-Othmer. *Encyclopaedia of Chemical Technology,* 4th edn. Wiley, 2, 1992, pp. 729–748.

Sternbach, R.A. *Pain Patients: Traits and Treatment*. Academic Press, New York, 1974.

Anesthetics

An anesthetic is a drug that causes total loss of sensory perception and thus enables complex surgical procedures to be carried out painlessly. The first surgical uses of ether and chloroform as inhalation anesthetics began in the 1840s, but they were administered crudely by holding a sponge or cloth saturated with the drug over the patient's nose and mouth. By 1910, mixtures of nitrous oxide with ether had begun to replace chloroform, though anesthetic ether and chloroform are still manufactured. Other compounds were soon introduced; and as surgeons demanded deeper, more controlled levels of anesthesia, the professional anesthetist became an essential member of every surgical team.

There are many factors to be considered in choosing a suitable anesthetic agent for each patient. The ideal anesthetic should allow rapid induction followed by ease of control and the possibility of rapid reversal. It should give good muscle relaxation, have few toxic or adverse effects, and be stable in soda lime, used in anesthesia equipment to absorb expired carbon dioxide. Inhalation anesthetics are administered in mixtures with oxygen alone or with a 30/70 mixture of nitrous oxide and oxygen. They are also often combined with drugs that relax muscles and block neuromuscular impulse transmission, making surgical operations easier. Artificial respiration may be required to maintain proper levels of oxygen and carbon dioxide in the blood as deep anesthesia can bring the patient close to death.

Most modern inhalation anesthetics are synthetic halogen-substituted hydrocarbons or ethers. Fluorine, the most common halogen substitute, decreases flammability and boiling point and increases the stability of the molecule. It also reduces fluctuations in heart rate. Fluroxene, introduced in 1960, gives rapid onset and recovery and is stable in soda lime; however, it is unstable in light and readily metabolized in the body. It was replaced in 1962 by methoxyflurane and in 1963 by enflurane. Methoxyflurane is the most potent of the inhaled anesthetics, but like fluroxene it is metabolized. Because fluoride ions cause renal damage, the duration of methoxyflurane anesthesia must be limited. Enflurane is the least potent of the inhaled anesthetics. Later inhalation anesthetics include Sevoflurane, launched in Japan in 1989, and Desflurane. Both are less subject to metabolism in

the body, but Sevoflurane is unstable in soda lime. Modern research has been concentrated on the discovery of safer anesthetic agents, and new compounds steadily replace older, less satisfactory ones. However, all halogen-substituted anesthetics tend to trigger hypermetabolic reactions accompanied by rapid temperature rise, increased oxygen use, and carbon dioxide production.

In addition to inhalation techniques, anesthesia may also be produced by intravenous injection. There are two types of injection anesthetics, those used only to induce anesthesia and those used both to induce and maintain anesthesia. Since the ideal injection agent is yet to be discovered, a ''balanced'' anesthesia is frequently employed with one drug injected for rapid induction followed by an inhalation agent to maintain anesthesia.

French surgeon Pierre Cyrpien Oré first attempted to produce anesthesia by intravenous injection of chloral hydrate in 1874. Improved safety was achieved with hedonal (methylpropylurethane) in 1899, and this was followed during the early 1900s by the intravenous injection of diluted chloroform and ether, as well as hedonal. Toward the end of World War I, barbiturates manufactured by the Swiss firm Hofmann–La Roche were introduced. In the 1920s chemical research revealed a large number of barbiturates, including thiopental sodium (Pentothal) and hexobarbitone, which is said to have been used in up to 10 million cases by the end of World War II. Pentothal is still widely used, but another group of drugs, the benzodiazepines, including diazepam (Valium) and chlordiazepoxide hydrochloride (Librium), were introduced in the early 1960s as muscle relaxants and tranquilizers. Several other drugs in this range are used in neurosurgery and cardiovascular surgery. Among the opiates, morphine is the most common and most potent drug. It is used in high doses as an anesthetic, despite its propensity to cause nausea and vomiting, respiratory and cardiovascular depression, and hypotension (lowered blood pressure). Fentanyl citrate, a synthetic opioid 50 to 100 times more potent than morphine, was introduced in the 1960s. It is used in cardiac surgery in very large doses to produce profound analgesia and suppress cardiovascular reflexes. Since the 1980s other synthetic opioids have been approved for clinical use.

An important difference between inhalation and injection anesthetics is that the former exert physical effects on the respiratory system, whereas the latter function by combining chemically with receptor molecules in the cells. There are two kinds of receptors, the GABA (γ-aminobutyric acid) receptor and the opiate receptor. As the injected anesthetic is chemically bound to its receptor, removal from the system is slow and other drugs (antagonists) are required to reverse the anesthetic effects.

A third type, local anesthetics, produces loss of sensation in limited areas without loss of consciousness. They are usually administered by subcutaneous injection around sensory nerve endings to block the passage of nervous impulses. Some local anesthetics also block the motor nerves enabling operations to be carried out while the patient remains conscious. The choice of local anesthetic depends on the type and length of the dental or surgical procedure for which it is to be used. The ideal local anesthetic should have rapid onset and long duration, and it should be useful in situations requiring differential blockage of sensory and motor nerve fibers. Obstetric and postoperative pain relief requires a powerful sensory block accompanied by minimal motor block, whereas limb surgery requires both sensory and motor block. A special form of regional nerve block may be caused by injecting a local anesthetic into the spinal cord, either into the space between the membranes that surround the cord (epidural anesthesia) or into the cerebrospinal fluid (spinal anesthesia). In these cases the anesthetic can be adjusted to block conduction in nerves entering and leaving the cord at the desired level.

New compounds are constantly being investigated for anesthetic properties, though the anesthetist now has a wide range of choice. It seems likely that future advances in anesthesia will depend upon developments in the computerization of monitoring and control of the patient's physiological status in response to these agents rather than on the discovery of new anesthetic compounds.

NOEL G. COLEY

Further Reading

Adams, R.C. *Intravenous Anesthesia*. Hoeber, New York, 1944.

Armstrong Davison, M.H. *The Evolution of Anesthesia*, revised edn. John Sherrat, Altrincham, 1970.

Atkinson, R.S., Rushman, G.B. and Lee J.A. *A Synopsis of Anesthesia*, 10th edn. IOP Publications, Bristol, 1987. Surveys late twentieth century anesthesia and analgesia practices.

Duncum, B.M. *The Development of Inhalation Anesthesia*. Royal Society of Medicine Press, London, 1994

Eriksson, E. *Illustrated Handbook in Local Anesthesia*. W.B. Saunders, Philadelphia, 1979.

Holden, A.V. and Winlow, W. *The Neurobiology of Pain*. Manchester University Press, Manchester, 1984.

Kirk-Othmer. *Encyclopaedia of Chemical Technology,* 4th edn. Wiley, New York, 1992, 2, 778–800.
Smith, E.B. and Daniels, S. Eds. *Gases in Medicine: Anesthesia*. Proceedings of the 8th BOC Priestley Conference, London, Royal Society of Chemistry, 1998.

Angiography

Angiography is a radiographic procedure that employs x-rays of the blood vessels in the body to assist in diagnosing and treating a variety of ailments. The x-ray image, or angiogram, allows diagnosis of pathologies of the blood vessels such as blockages, stenosis (narrowing of the vessel), and other aberrations so that they can be treated. Angiography can be purely investigative, but it is commonly employed in combination with minimally invasive surgery or catheterization.

In the angiography procedure, a venous "contrast," or dye, is administered orally, anally, or by intravenous injection. This dye is a radiopaque substance that highlights the blood vessels in the x-ray.

Magnetic resonance angiography (MRA) does not rely on radioactive dye; rather it employs a specific sequence of radio waves to create the angiogram. The MRA is able to provide a detailed map of the patient's vasculature without an enhancer, although enhancers such as the rare earth element gadolinium are sometimes used to make the images clearer and bolder. In therapeutic or interventional angiography, the possible treatment runs from surgery to less invasive processes such as angioplasty or catheterization. In angioplasty, a catheter containing a small balloon is guided through the blood vessels to the site of the obstruction. Once in place, the balloon is inflated in order to expand the constricted area of the vessel. The catheter can also be used to guide into place surgical "stents," cylindrical mesh-like supports that keep the vessel open to the desired width. Over one million angioplasty procedures were performed in the U.S. in 2000 to prevent or treat myocardial infarction (heart attack).

Rudimentary angiography was developed not long after the x-ray came into clinical use. In 1896 in Vienna, Eduard Haschek and Otto Lindenthal took x-rays of the blood vessels in an amputated hand injected with a radiopaque substance. The radiopaque contrast agent of choice has changed over time, owing in part to the high toxicity of the earliest agents used. Common formulations of contrast agents used today include various acetrizoic acids, diatrizoic acids, iodamides, and methanesulfonic acids.

Angiography is employed today in a number of clinical situations. The most common procedures performed are cerebral angiography, thoracic aortography, pulmonary and bronchial arteriography, coronary arteriography, and angiography of the extremities. In celial and mesenteric angiography, these arteries in the abdomen are examined to diagnose gastrointestinal bleeding, aneurysm, or ischemia. In cerebral angiography the blood vessels of the brain are investigated in order to locate and treat of blood clots, aneurysms, tumors, and migraines. In coronary angiography the coronary arteries are investigated to detect vascular diseases such as heart disease, heart attack, and acute stroke. Renal angiography is of value in diagnosing kidney disease and renal failure. In the latter case, the physician uses hemodialysis catheters to divert blood from the neck and filter it through a hemodialysis machine. One problem with kidney angiography is that the contrast agent can be harmful to the kidneys if the patient already suffers from renal ailments. In ocular angiography either fluorescein or indocyanine green contrast helps to view various problems in the eye such as retinal failure.

Complications surrounding angiographic procedures can arise either from reactions to the contrast agent or from problems with the performance of the catheterization or surgery. Reactions to the contrast agent are generally mild, such as nausea, but serious allergic reactions do occasionally occur. Renal damage can occur regardless of previous problems because of the mild toxicity of most opaque media contrasts. Some contrasts have also caused minor drops in blood pressure and vasodilatation of the arteries. When a catheter is placed in the body, blood clots can form and block the vessel or a small particle can break loose and lead to embolization, a potentially deadly complication. Catheters can also tear or puncture blood vessels causing internal bleeding and an exacerbation of the already existing problem. Hematomas and hemorrhages may occur if there are any complications in the catheterization process.

As with other medical tests and procedures, the risks of angiography are generally outweighed by the benefits, as the angiogram provides specific and detailed clinical information that is invaluable in clinical diagnosis and intervention.

See also **Cardiovascular Disease, Diagnostic Methods; Medicine; Nuclear Magnetic Resonance (NMR/MRI); X-rays in Diagnostic Medicine**

TIMOTHY S. BROWN

Further Reading

Abrams, H.L. *Angiography*. Little, Brown & Company, Boston, 1971.
Grossman, W. and Bain, D.S. *Cardiac Catheterization, Angiography and Intervention*. Lea & Febiger, Philadelphia, 1991.
Silverman, J.F. *Coronary Angiography: An Introduction to Interpretation and Technique*. Addison-Wesley, Menlo Park, CA, 1984.
Tortorici, M.R. *Advanced Radiographic and Angiographic Procedures*. F.A. Davis, Philadelphia, 1995.

Antibacterial Chemotherapy

In the early years of the twentieth century, the search for agents that would be effective against internal infections proceeded along two main routes. The first was a search for naturally occurring substances that were effective against microorganisms (antibiosis). The second was a search for chemicals that would have the same effect (chemotherapy).

Despite the success of penicillin in the 1940s, the major early advances in the treatment of infection occurred not through antibiosis but through chemotherapy. The principle behind chemotherapy was that there was a relationship between chemical structure and pharmacological action. The founder of this concept was Paul Erhlich (1854–1915). An early success came in 1905 when atoxyl (an organic arsenic compound) was shown to destroy trypanosomes, the microbes that caused sleeping sickness. Unfortunately, atoxyl also damaged the optic nerve. Subsequently, Erhlich and his co-workers synthesized and tested hundreds of related arsenic compounds. Ehrlich was a co-recipient (with Ilya Ilyich Mechnikov) of the Nobel Prize in medicine in 1908 for his work on immunity.

Clinical trials in 1910 showed that another of these compounds was effective in treating not only sleeping sickness but also syphilis. It was called arsphenamine and given the trade name Salvarsan. The success of Salvarsan was evidence that synthetic drugs could destroy microbes in patients, but again there were severe limitations. A course of treatment involved an injection a week for 10 to 12 weeks. This had to be repeated two or three times, with intervals of one month or so to minimize toxic effects. Treating early syphilis took a year or more.

Some progress in antibacterial chemotherapy was made between 1920 and 1932 in the laboratories of the Bayer Company in Elberfeld, Germany. Ehrlich had developed simple methods for mass screening. He selected a suitable test system (typically a group of animals infected with a particular organism) and used it to test many substances, most of which were newly synthesized. But real progress occurred only after the appointment by Bayer of Gerhard Domagk as director of research in experimental pathology in 1927. In collaboration with two chemists, Fritz Mietzsch and Joseph Klarer, Domagk immediately began a program to develop antibacterial drugs.

Domagk and his team investigated dyestuffs as a possible source of potential drugs. In 1932 they experimented on two groups of mice, all of which had been infected with streptococci. One group was treated with a red dye called prontosil, and the other was not. The untreated group all died within four days; the treated group all survived at least a week. For commercial reasons nothing was published for over two years, during which time Domagk treated his own daughter successfully for a streptococcal infection. By 1935 a large number of clinical trials were under way, and the drug rapidly became widely available. Domagk received the Nobel Prize in medicine in 1939.

Later in 1935 Trefouel and his colleagues at the Pasteur Institute in Paris showed that prontosil was broken down *in vivo* (in the body) to a much simpler compound, sulfanilamide, which was the active disease-fighting agent. Prontosil is not active *in vitro* (outside the body), and this nearly led to the failure of its discovery. The two chemists Mietzsch and Klarer were uneasy about performing initial tests in mice and wanted to screen compounds on bacterial culture in test tubes first. If this had been done, the action of prontosil would have been missed, not for the first but for a *second* time, because sulfanilamide was not a new substance. It had been synthesized and described in a publication in 1908 as part of research on dyes. However, a major obstacle to progress in the development of antibiotics in the 1920s was the prevalence of contemporary ideas. Sulfanilamide had actually been tested as an antibacterial in about 1921 by Michael Heidelberger at the Rockefeller Institute in New York, but he failed to discover its activity. It was believed at that time that if an antibacterial drug were to be effective at all, it would work immediately. In fact, sulfanilamide takes several hours to produce its effects, and Heidelberger stood no chance of observing them under the conditions he used.

It took some time to work out the mode of action of sulfanilamide. It was a powerful and effective drug, but it did not work under certain conditions; for example, when the target streptococci were surrounded by dead tissue or pus. Eventually the process of *competitive antagonism* was understood. Microbes that are sensitive to

sulfanilamide need a substance known as *p*-aminobenzoic acid (or PABA) as a building block for their growth. Sulfanilamide is a false building block shaped like PABA and is able to take its place. Once in position it prevents any further building and cell growth ends. The cells are then dealt with by the body's normal defensive mechanisms, and the patient recovers.

Soon after this discovery, a large number of related compounds, collectively known as the sulfonamides, were synthesized by modification of the parent compound. They were shown to be effective against various bacterial infections. Sulfanilamide was very effective against streptococcal infections, including puerperal fever, erysipelas, and scarlet fever. It was, however, not active against pneumococci, which cause pneumonia. A pharmaceutical company in the U.K. (May and Baker) soon developed a sulfonamide that was. Called sulfapyridine, it was marketed as M&B 693. The first trials were carried out at the Middlesex Hospital in London in 1938, under the supervision of Sir Lionel Whitby.

Other sulfonamides were being developed during this time. Trials indicating the range of antibacterial activity of sulfathiazole were reported in 1940. Sulfadiazine followed in 1941 after trials on 446 adult patients at Boston City Hospital. In 1942 the results of early trials on a sulfonamide developed by Imperial Chemical Industries, sulfadimidine, were published, indicating its value in the treatment of pneumonia, meningitis, and gonorrhoea. Sulfaguanidine was used successfully during World War II to treat bacillary dysentery in troops in the Middle and Far East.

Success in discovering a range of effective antibacterial drugs had three important consequences: it brought a range of important diseases under control for the first time; it provided a tremendous stimulus to research workers and opened up new avenues of research; and in the resulting commercial optimism, it led to heavy postwar investment in the pharmaceutical industry. The therapeutic revolution had begun.

See also **Antibiotics: Developments through 1945; Antibiotics: Use after 1945.**

STUART ANDERSON

Further Reading

Bickel, M.H. The development of sulphonamides, 1932–1938. *Gesnerus*, 45, 67–86, 1988.

Lesch, J. The Invention of the Sulfa Drugs, in *Chemical Sciences in the Modern World,* Mauskopf, S.H., Ed. University of Pennsylvania Press, Philadelphia, 1993.

Lesch, J. The Discovery of M&B 693 (Sulfapyridine), in *The Inside Story of Medicines: A Symposium,* Higby, G. and Stroud, E., Eds. American Institute of the History of Pharmacy, Madison, 1997.

Liebenau, J., Higby, G. and Stroud, E., Eds. *Pill Peddlers: Essays on the History of the Pharmaceutical Industry*. American Institute of the History of Pharmacy, Madison, 1990.

Mann, R.D. *Modern Drug Use: An Enquiry on Historical Principles*. MTP Press, the Hague, 556–57, 1984.

Slinn, J. *A History of May and Baker 1834–1984*. Hobsons, Cambridge, 1984.

Tweedale, G. *At the Sign of the Plough: 275 Years of Allen & Hanburys and the British Pharmaceutical Industry 1715–1990*. John Murray, London, 1990.

Weatherall, M. *In Search of a Cure: A History of Pharmaceutical Discovery*. Oxford University Press, Oxford, 1990.

Antibiotics, Developments through 1945

The term antibiotic was first used in 1899 by the botanist Marshall Hall, following the coining of the term antibiosis ten years earlier by French scientist Paul Vuillemin. Antibiosis referred to the idea that living organisms might produce substances that were antagonistic to one another, an idea first suggested by Louis Pasteur in 1877. In the late nineteenth century, many observations of antibiotic action among microorganisms were reported, but unfortunately the significance of these observations was not appreciated at the time.

In the history of antibiosis, an observation by Alexander Fleming (1881–1955), a bacteriologist at St. Mary's Hospital in London, proved to be the turning point. In the late summer of 1928, Fleming went to his laboratory during the holiday to inspect culture plates awaiting disposal. He noticed one that had been accidentally contaminated with a mold, or fungus, around which bacterial growth had been inhibited. The bacteria were staphylococci and the fungus was *Penicillium notatum*. Fleming cultured it, fixed the plate in formalin vapor, and gave it the name penicillin.

Fleming's tests showed that the fungus could kill or inhibit the growth of a number of other organisms harmful to man. In a paper in 1929, Fleming mentioned the possible use of penicillin as an antiseptic in surgical dressings. There was further work in other laboratories interested in Fleming's observation, but progress was limited because of the great instability of the material. The goal of isolating an effective preparation from the organism remained for the future.

In 1939, three other British scientists based in Oxford made the discovery that penicillin was

effective inside the body as well as on its surface. Ernst Chain, Howard Florey, and Norman Heatley, embarked on its detailed investigation, urged on by the necessities of war. In 1940 Florey and Chain were able to make a dry, stable extract of penicillin, and the first trial was carried out on four mice on May 25 1940.

The first human received an injection of penicillin on January 27 1941, at the Radcliffe Infirmary in Oxford. The patient did not have an infection, but suffered a sharp rise in temperature after 3 hours. After removal of the pyrogen that caused the fever, enough active material was isolated to treat one ill patient. On February 12 1941, a London policeman who had a severe staphylococcal infection was the first sick patient to receive penicillin. After treatment for five days, the stock of penicillin was exhausted, and the patient relapsed and died. Over the following days further seriously ill patients were treated with varying degrees of success, and the results were published in August 1941.

By early 1942, small-scale production was being carried out in the U.K. by Imperial Chemical Industries Limited. However the investment needed to produce penicillin on a commercial scale was considerable and could not easily be found in a country at war. Florey and Heatley went to America to seek support. Penicillin was first made in the United States by the pharmaceutical company Upjohn in March 1942, using a culture supplied by the U.S. Department of Agriculture. Initial small-scale production was achieved using a surface culture method in bottles. On May 25 1943, the firm was asked to make penicillin for the military, and it began brewing the fungus in 120,000 bottles in a basement. By July 1943 the company had shipped its first batch of 100 vials, each containing 100,000 units of penicillin.

The process of surface culture in bottles had obvious limitations, and work began to develop a more efficient process that used deep culture in vat fermenters. The principle of the submerged culture method is to grow the fungus in large steel containers in a medium that is constantly aerated and agitated. Under these circumstances, the fungus grows throughout the body of the medium. Early problems in the operation of the method were overcome, and by 1949 fermenters with capacities up to 12,000 gallons were in use.

Tens of thousands of species and mutants of *Penicillium* were examined in the search for ones that produced the highest yields of penicillin. Eventually the highest yields were obtained from a strain of *Penicillium chrysogenum,* originally found growing naturally on a moldy melon. As a result yields have increased 2000 times since the original small-scale manufacture.

Solving the complex microbiological, chemical, and engineering problems involved in the large-scale manufacture of penicillin required a collaborative effort. By 1944, 20 American and 16 British academic and industrial groups were working on the problem. The first large delivery, consisting of 550 vials of 100,000 units of penicillin each, was made to the U.S. Army in February 1944. Its initial use was in the treatment of gonorrhea; widespread use of penicillin in the treatment of life-threatening conditions such as pneumonia only occurred later when sufficient supplies became available. With large-scale manufacture the price dropped dramatically from \$20.05 per 100,000-unit vial in July 1943 to just \$0.60 per vial by October 1945.

Full understanding of the mode of action of penicillin only emerged slowly and awaited developments in both biochemistry and bacteriology. The involvement of the cell wall was recognized early on, but even in 1949 Chain and Florey concluded, "No complete picture of how penicillin acts *in vivo* can be drawn on the evidence available."

Penicillin inhibits an enzyme that catalyzes one of the biochemical steps in the synthesis of mucopeptide, the rigid component of the cell wall. In the absence of the enzyme inhibited by penicillin, links between the peptide molecules fail to occur. When penicillin in sufficient concentration comes into contact with penicillin-sensitive organisms, it binds to the cell wall. As cell division takes place, defects occur in the rigid component of the wall. Coupled with high internal osmotic pressure, the cell then bursts, or lyses. Penicillin is therefore only active against growing bacteria and only those sensitive to it. Some bacteria contain plasmids that make them resistant to penicillin.

In subsequent years a large number of variants of penicillin were developed. Some of these were naturally produced, some were semisynthetic, and others were entirely synthetic. The search for effective antibacterial substances was widened, and the stage was set for the discovery of a wide range of new antibacterial substances in the post-World War II years.

See also **Antibacterial Chemotherapy; Antibiotics: Use after 1945**

STUART ANDERSON

Further Reading

Carlyle, R.D.B. *A Century of Caring: The Upjohn Story*. Benjamin, New York, 1987, pp. 89–95.

Elder, A., Ed. *The History of Penicillin Production*. Livingstone, New York, 1970.

Florey, H.W., Chain, E., Heatley, N.G., Jennings, M.A., Sanders, A.G., Abraham, E.P. and Flory, M.E. *Antibiotics: A Survey of Penicillin, Streptomycin, and other Antimicrobial Substances from Fungi, Actinomycetes, Bacteria and Plants*. Oxford University Press, Oxford, 1949, 1, pp. 1–73.

Florey, H.W., Chain, E., Heatley, N.G., Jennings, M.A., Sanders, A.G., Abraham, E.P. and Flory, M.E. *Antibiotics: A Survey of Penicillin, Streptomycin, and other Antimicrobial Substances from Fungi, Actinomycetes, Bacteria and Plants*. Oxford University Press, Oxford, 1949, 2, pp. 631–671.

Hare, R. *The Birth of Penicillin*. Allen & Unwin, London, 1970

Lazell, H.G. *From Pills to Penicillin: The Beecham Story*. Heinemann, London, 1975.

Maurois, A. *The Life of Alexander Fleming: Discoverer of Penicillin,* translated by Hopkins, G. Heinemann, London, 1952.

Weatherall, M. *In Search of a Cure: A History of Pharmaceutical Discovery*. Oxford University Press, Oxford, 1990.

Antibiotics: Use after 1945

Antibiotics, a category of antiinfective drugs, are chemical substances produced by microorganisms such as bacteria and fungi that at low concentrations kill or inhibit other microorganisms. The development of antibiotic drugs was one of the most important advances in twentieth century medicine. Penicillin, the first antibiotic, was first produced in the U.S. in 1942. About the same time new antibiotics, including streptomycin, were discovered in a mixture of microbial substances from a soil bacillus. Many of these drugs were found to be specific in their action but in 1947, chloramphenicol, the first so-called broad spectrum antibiotic was discovered; followed in 1948 by aureomycin, the first broad spectrum tetracycline antibiotic. Chloramphenicol, a nitrobenzene derivative of dichloroacetic acid, is manufactured synthetically and used to treat typhoid fever and severe infections caused by penicillin-resistant microorganisms.

As antibiotics destroyed the more common pathogens, highly resistant strains that escape destruction began to pose problems, and new antibiotics were required to eradicate them. By the end of the twentieth century, there were well over 10,000 known antibiotics ranging from simple to extremely complex substances. This great expansion in new antibiotics was achieved through successful synthetic modifications, and many antibiotics are now manufactured in quantity. This approach has been especially successful with antibiotics whose molecular structure contains the four-member β-lactam ring, including the penicillins and cephalosporins that currently account for over 60 percent of world production of antibiotics (see Table 1 for major types of antibiotics).

The modes of action of different antibiotics vary, though many destroy bacteria by inhibiting cell-wall biosynthesis. Bacterial cells, unlike animal cells, have a cell wall surrounding a cytoplasmic membrane. The cell wall is produced from components within the cell that are transported through the cell membrane and built into the cell wall, and antibiotics that interfere with this process inhibit cell growth. For example, the β-lactam ring in penicillin and cephalosporin molecules interferes with the final stage in the assembly of the cell wall. Other antibiotics inhibit protein synthesis in bacteria, while others act by combining with phos-

Table 1 Important Groups of Antibiotics.

Classification	**Generic Drug**
Aminoglycoside	Gentamicin sulfate
Antifungal	Fluconazole
β-Lactam	Imipenim-Cilastatin sodium
Cephalosporin	
First Generation	Cephalothin sodium
Second Generation	Cefonicid sodium
Third Generation	Cefotaxime sodium
Clindamycin	Clindamycin hydrochloride
Macrolide	Erythromycin
Penicillin	
Aminopenicillin	Ampicillin
Antipseudomonal Penicillin	Mezlocillin sodium
Natural Penicillin	Penicillin G potassium
Tetracycline	Tetracycline hydrochloride

pholipids in the cell membrane to interfere with its function as a selective barrier and allow essential materials in the cell to leak out. This results in the death of the bacterial cell; but since similar phospholipids are found in mammalian cells, this type of antibiotic is toxic and must be used with care. One antibiotic, rifampin, disrupts RNA synthesis in bacterial cells by combining with enzymes in the cell; as its affinity for bacterial enzymes is greater than for mammalian ones, the latter remain largely unaffected at therapeutic dosages.

Penicillin is the safest of all antibiotics, although some patients show hypersensitivity toward it, and this results in adverse reactions. Moreover some microorganisms, notably the staphylococci, develop resistance to naturally occurring penicillin, and this has led to the production of new synthetic modifications. Thus there are two groups of penicillins, those that occur naturally and the semisynthetic penicillins made by growing the *Penicillium* mold in the presence of certain chemicals. To increase the usefulness of the penicillins, the broad-spectrum penicillins were developed to treat typhoid and enteric fevers and certain infections of the urinary tract. Nevertheless, the naturally occurring penicillins are still the drugs chosen for treating many bacterial infections.

The cephalosporins, discovered in the 1950s by Sir Edward Abraham, are relatively nontoxic β-lactam antibiotics. Like penicillin they were first isolated from a fungus, but later modifications of the β-lactam ring have resulted in over 20 variations grouped according to their activity. The first generation cephalosporins were used for patients who had developed sensitivity to penicillin. They were active against many bacteria including *Escherichia coli*, but they had to be replaced by second and third generation drugs as resistance began to develop. They have been used for treating pulmonary infections, gonorrhea, and meningitis.

The aminoglycosides, which include streptomycin (discovered in 1944), inhibit protein biosynthesis. They are poorly absorbed from the gastrointestinal tract and are administered by intramuscular injection. Streptomycin was among the first of the aminoglycosides to be discovered and is still used together with penicillin for treating infections of the heart valves. Other aminoglycosides are used for treating meningitis, septicemia, and urinary tract infections, but the narrow margin between a therapeutic and a toxic dose poses difficult problems, and the risks increase with age. The antimicrobial activity of the tetracyclines, another group of synthetic antibiotic drugs, depends on the fact that although they inhibit protein biosynthesis in both bacterial and animal cells, bacteria allow the tetracyclines to penetrate the cell, whereas animal cells do not. Tetracyclines are absorbed from the gastrointestinal tract and can be given orally.

In addition to their uses in medicine, antibiotics have also had important veterinary applications and are used as animal feed supplements to promote growth in livestock. Tetracyclines make up about half the sales of antibiotics as supplements, which surpasses all other agricultural applications, but many other antibiotics are also used for this purpose. It is thought that feed antibiotics may promote growth by preventing disease. Another important agricultural use of antibiotics is in their use as antiparasitic agents against both worms and other parasites in the gastrointestinal tract and against ectoparasites such as mites and ticks.

In addition to the biochemical antibiotics, the sulfonamides, synthetic chemotherapeutic agents, are also used in treating bacterial diseases. The first sulfonamide was prontosil, introduced in 1932 to combat streptococcal infections. The sulfonamides are broad-spectrum agents that were widely used prior to antibiotics. They work by preventing the production of folic acid, which is essential for the synthesis of nucleic acids. The reaction is reversible causing the inhibition but not the death of the microorganisms involved. Their use has diminished due to the availability of better and safer antibiotics, but they are still used effectively to treat urinary tract infections and malaria and to prevent infection after burns. Related to the sulfonamides, the sulfones are also inhibitors of folic acid biosynthesis. They tend to accumulate in the skin and inflamed tissue, and as they are retained in the tissue for long periods, they are useful in treating leprosy. There are also some other chemical synthesized drugs that show antibacterial properties and find specific clinical uses.

See also **Antibacterial Chemotherapy; Antibiotics: Developments through 1945**

NOEL G. COLEY

Further Reading

Bottcher, H. *Miracle Drugs: a History of Antibiotics*. Heinemann, London, 1959.

Cannon, G. Superbug: *Nature's Revenge. Why Antibiotics Can Breed Disease*. Virgin, London, 1995.

Halliday, J., Ed. *Antibiotics: The Comprehensive Guide*. Bloomsbury, London, 1990.

Kirk-Othmer. *Encyclopaedia of Chemical Technology*, 4th edn. Wiley, New York, 1992, 2, pp. 893–1018 and 3, pp. 1–346.
MacFarlane, M. *Alexander Fleming: The Man and the Myth*. Chatto and Windus, Hogarth Press, London, 1984.
Pharmaceutical Society of Great Britain. *Antibiotics: A Survey of Their Properties and Uses*. Pharmaceutical Press, London, 1952.
Reese R.E., Betts R.F. and Gumustop B. *Handbook of Antibiotics*, 3rd edn. Lippincott, Williams & Wilkins, New York, 2000.

Architecture, *see* **Constructed World**

Artificial Insemination and *In Vitro* Fertilization (Farming)

Artificial insemination (AI) involves the extraction and collection of semen together with techniques for depositing semen in the uterus in order to achieve successful fertilization and pregnancy. Throughout the twentieth century, the approach has offered animal breeders the advantage of being able to utilize the best available breeding stock and at the correct time within the female reproductive cycle, but without the limitations of having the animals in the same location. AI has been applied most intensively within the dairy and beef cattle industries and to a lesser extent horse breeding and numerous other domesticated species.

There is some anecdotal historical evidence to indicate the ancient use of AI in mammalian reproduction, dating as far back as fourteenth century Arabia, telling of how rival tribes would obtain sperm from the mated mares or from the male horses of their opponents and use it to inseminate their own mares. Based on the first visible identification of spermatozoa in the seventeenth century, the earliest documented insemination is believed to have been carried out by Italian physiologist Lazzaro Spallanzani in 1784 on a domestic dog, resulting in a successful pregnancy. This was followed a century later with the work in the U.K. by Cambridge reproductive scientist Walter Heape who in 1897 documented the importance and variation of reproductive cycles and seasonality across a number of species. However, it was not until the opening decades of the twentieth century with the work of Russian reproductive scientist E. I. Ivanow that AI was more widely investigated, particularly in Scandinavia, though still largely within the confines of research rather than general animal husbandry.

The emergence of an AI research community takes firmer shape with the first International Congress on AI and Animal Reproduction in Milan in 1948. The 1940s can be understood as the decade in which the technique progresses from a relatively small-scale research activity to becoming one of the most routine of procedures in domestic animal reproduction. The establishment of the New York Artificial Breeders Cooperative in the early 1940s resulted in the insemination of hundreds of thousands of cattle and the refinement of many of the methods on which AI would subsequently come to depend. Progress on techniques relevant to male aspects of AI includes sire selection, semen collection, and evaluation methods, as well as techniques for safely preserving and storing semen. In relation to female aspects of AI, innovation has focused on characterizing species-specific patterns of estrus detection and synchronization in order to best judge the timing of insemination. As the twentieth century drew to a close, in addition to becoming one of the most widely used techniques within animal breeding, AI has become ever more prominent in the preservation of rare and endangered species. In 1986, the American Zoo and Aquarium Association established a Species Survival Plan, which has been successful in reviving such nearly extinct animals as black-footed ferrets.

In Vitro *Fertilization*

Many of the techniques involved in artificial insemination would lay the foundation for *in vitro* fertilization (IVF) in the latter half of the twentieth century. IVF refers to the group of technologies that allow fertilization to take place outside the body involving the retrieval of ova or eggs from the female and sperm from the male, which are then combined in artificial, or "test tube," conditions leading to fertilization. The fertilized eggs then continue to develop for several days "in culture" until being transferred to the female recipient to continue developing within the uterus.

The first reported attempts to fertilize mammalian ova outside the body date back as far as 1878 and the reported attempt to fertilize guinea pig eggs on a solution of mucus taken from the uterus. However, the first verified successful application of IVF did not occur until 1959 and resulted in the live derivation of rabbit offspring. This was followed in the late 1960s with mice and coincided with similar but unsuccessful attempts at IVF in humans by Robert Edwards, Barry Bavister, and others. However, it was events in 1978 that decisively put the technique on the research agenda

for reproductive science in both humans and other mammals. The birth of Louise Brown in the U.K. in 1978, the first successful application of the technique in humans, represented the point at which IVF became an intense focus of reproductive activity throughout the latter years of the twentieth century. By the early twentieth century IVF was an established reproductive technology in the production of highly prized genetic source animals and, as with AI, the preservation of endangered species.

See also **Agriculture and Food; Fertility, Human**

NIK BROWN

Further Reading

Cupps, P.T. *Reproduction in Domestic Animals*, 4th edn. Academic Press, San Diego, 1991.

Foote, R.H. *The History of Artificial Insemination. Selected Notes and Notables*. American Society of Animal Science, 2002.

Foote, R.H. Artificial Insemination from its origins up to today, in Russon, V., Dall'Olio, S. and Fontanesi, L., Eds. *Proceedings of the Spallanzani International Symposium,* Reggio Emilia, Italy, 1999, pp 23–67.

Foote, R.H. The artificial insemination industry, in *New Technologies in Animal Breeding*, Brackett, B.G., Seidel, Jr., G.E. and Seidel, S.M., Eds. Academic Press, New York, 1981, pp 13–39.

Herman, H.A. *Improving Cattle by the Millions: NAAB and the Development and Worldwide Application of Artificial Insemination*. University of Missouri Press, Columbia, 1981.

Perry, E.J., Ed. *The Artificial Insemination of Farm Animals*, 4th edn. Rutgers University Press, New Brunswick, 1968.

Sipher, E. *The Gene Revolution: The History of Cooperative Artificial Breeding in New York and New England 1938–1940*. Eastern AI Cooperative, Ithica, 1991.

Trounson, A.O. *et al. Handbook of In Vitro Fertilization*. CRC Press, London, 2000.

Artificial Intelligence

Artificial intelligence (AI) is the field of software engineering that builds computer systems and occasionally robots to perform tasks that require intelligence. The term ''artificial intelligence'' was coined by John McCarthy in 1958, then a graduate student at Princeton, at a summer workshop held at Dartmouth in 1956. This two month workshop marks the official birth of AI, which brought together young researchers who would nurture the field as it grew over the next several decades: Marvin Minsky, Claude Shannon, Arthur Samuel, Ray Solomonoff, Oliver Selfridge, Allen Newell, and Herbert Simon.

During the 1940s many researchers, under the guise of cybernetics, had worked out much of the theoretical groundwork for AI and had even designed the first computers. Among the most significant contributions upon which AI depended were Alan Turing's theory of computation and ACE computer, John Von Neumann's ENIAC computer, Claude Shannon's theory of communication, Norbert Weiner's theory of information and negative feedback, Warren McCulloch and Walter Pitts' neuronal logic networks, W. Ross Ashby's theory of learning and adaptive mechanisms, and W. Grey Walter's autonomous robotic tortoises. The young Dartmouth group differed from the earlier work of the cyberneticians in that they concerned themselves primarily with writing digital computer programs that performed tasks deemed to require intelligence for humans, rather than building machines or modeling brains.

AI research focused around the key aspects of intelligent behavior including automated reasoning, decision making, machine learning, machine vision, natural language processing, pattern recognition, automated planning, problem solving, and robot control. This field of research set itself ambitious goals, seeking to build machines that could ''out-think'' humans in particular domains of skill and knowledge, and achieving some success in this. Some researchers even speculated that it would be possible to build machines that could imitate human behavior in general by the end of the twentieth century, but most researchers now consider this goal to be unattainable by the end of the twenty-first century.

The first AI program, Logic Theorist, was presented by Newell and Simon at the Dartmouth workshop in 1956. Logic Theorist proved theorems of mathematical logic from a given set of axioms and a set of rules for deducing new axioms from those it already had. Given a theorem, Logic Theorist would attempt to build a proof by trying various chains of deductive inference until it arrived at the desired theorem. Logic Theorist was followed by General Problem Solver in 1961. This program demonstrated that the technique of proving theorems could be applied to all sorts of problems by defining a ''goal'' and conducting a search to find a series of valid moves that led from what is already known to the goal that is sought. This technique can work well for simple problems, but the total number of alternative moves that are possible can grow exponentially in the number of steps to a solution. Since the program has to keep backtracking and trying all the alternate routes, the technique quickly breaks down for problems with many steps. These challenges led Newell and Simon to suggest that AI

research on problem solving ought to focus on finding good *heuristics,* search strategies or rules-of-thumb, to use when searching. A good heuristic helps one find a solution faster by reducing the number of dead ends encountered during a search.

On June 27, 1963, AI research was catapulted forward by a huge grant from the Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense to the Massachusetts Institute of Technology AI Laboratory. The grant was partly motivated by U.S. fears following the Soviet launch of Sputnik and partly motivated by extreme enthusiasm on the part of AI researchers that computers would soon have human-like intelligence. By the early 1970s, actual research at the Massachusetts Institute of Technology was limiting search spaces by studying very simplified application domains, or *micro worlds* as they came to be called. The most famous program, SHRDLU, planned manipulations in a world consisting only of wooden blocks sitting on a table called the *blocks world.* While these systems did what their designers intended, often led to theoretically interesting results, and eventually developed into useful technologies; they did not live up to either the public or military expectations for intelligent machines. By the end of the 1970s, DARPA became deeply concerned that AI would fail to deliver on its promises and eventually cut its research funding.

The insights gained from the micro worlds research eventually did find application in the area of automated planning. Planning generally starts from knowing the current state of the "world," the desired state of the world, and a set of actions called *operators* that can be taken to transform the world. The Stanford Research Institute Problem Solver (STRIPS) was an early planner that used a language to describe actions that is still widely used and enhanced. The STRIPS operators consist of three components:

1. The action description
2. The *preconditions* of the action, or the way the world must be before the action can be taken
3. The *effect*, or how the world has been changed since the action was taken. To develop a plan, the system searches for a reasonably short or cost-efficient sequence of operators that will achieve the goal. Planning systems have been used widely to generate production schedules in factories, to find the most efficient ways to lay out circuits on microchips or to machine metal parts, and to plan and coordinate complex projects involving many people and organizations, such as space shuttle launches.

One AI technology that had early real-world applications was the expert system. Expert systems utilize a large amount of knowledge about a small area of expertise in order to solve problems in that domain. The first such system was DENDRAL, which could logically infer the structure of a molecule if given its chemical formula and information from a mass spectrogram of the molecule. This difficult task was achieved by DENDRAL because it was provided with rules-of-thumb and tricks for recognizing common patterns in the spectrograms, developed in collaboration with Joshua Lederberg, a Nobel prize-winning chemist. The next generation expert system MYCIN used rules that incorporated uncertainty as probability weights on inferences. MYCIN used some 450 such rules to diagnose infectious blood diseases. Expert systems have proven to be one of the most successful applications of AI so far. Thousands of expert systems are currently in use for medical diagnoses, servicing and troubleshooting complex mechanical devices, and aiding information searches.

Another successful AI technology has been machine learning, which develops techniques for machines to actually learn from experience and improve over time. Machine learning was first conceived by Ashby in 1940, while the first successful program for learning was Samuel's 1959 checkers-playing program. Most forms of machine learning use statistical induction techniques to infer rules and discover relationships in sample or training data. Machine learning is useful in solving problems in which the rules governing the domain are difficult to discover and a large amount of data is available for analysis.

Pattern recognition is the most common type of problem for machine learning applications. The most popular pattern recognition systems, and perhaps the most popular single AI technology, are neural networks that learn from experience by adjusting weighted connections in a network. A typical feed-forward neural network performs some version of statistical pattern classification; that is, it induces statistical patterns from training data to learn a representative function, and then applies this function to classify future examples. A classification is simply a mapping function from inputs to outputs, and so a neural network just maps the objects to be classified into their types or classes.

Consider, for example, the problem of classifying some two-dimensional geometric shapes into one type of the set (square, circle, triangle, other). A total mapping function would assign every member of the set to one of these four types. There are many possible mapping functions however, and only a few of these will classify the inputs in a desirable way. Good techniques for neural networks will find these mappings efficiently and avoid getting stuck in statistical dead ends, called "local minima." Other examples of pattern recognition include speech recognition, face recognition, handwritten letter recognition, and robotic vision and scene analysis, in which the program must match audio or visual patterns to words, faces, letters, objects, or scenes, respectively.

Another important area of AI research has been natural language processing (NLP). NLP attempts to provide computers with the ability to understand natural human languages, such as English or Russian. Work in this area has drawn heavily on theories of grammar and syntax borrowed from computational linguistics and has attempted to decompose sentences into their grammatical structures, assign the correct meanings to each word, and interpret the overall meaning of the sentence. This task turned out to be very difficult because of the possible variations of language and the many kinds of ambiguity that exist. The applications of successful NLP programs have included machine translation from one natural language to another and natural language computer interfaces. A great deal of success has been achieved in the related areas of optical character recognition and speech recognition, which employ machine learning techniques to translate text and sound inputs into words but stop short of interpreting the meaning of those words.

Game playing programs have done much to popularize AI. Programs to play simple games like tic-tack-toe (noughts and crosses) are trivial, but games such as checkers (draughts) and chess are more difficult. At IBM, Samuel began working in 1952 on the program that would be the first to play tournament level checkers, a feat it achieved by learning from its own mistakes. The first computer to beat a human grandmaster in a chess match was HITECH in 1989. And in May of 1997, IBM's Deep Blue computer beat the top-ranked chess player in the world, Gary Kasparov. Unfortunately, success in a single domain such as chess does not translate into general intelligence, but it does demonstrate that seemingly intelligent behaviors can be automated at a level of performance that exceeds human capabilities. One of the most common consumer AI applications is probably the computer opponent in video games that "plays against" the human user. These applications use techniques of varying sophistication to challenge human opponents and often allow the human to select the skill level of their opponent.

It would be difficult to argue that the technologies derived from AI research had a profound effect on our way of life by the end of the twentieth century. However, AI technologies have been successfully applied in many industrial settings, medicine and health care, and video games. Programming techniques developed in AI research were incorporated into more widespread programming practices, such as high-level programming languages and time-sharing operating systems. While AI did not succeed in constructing a computer which displays the general mental capabilities of a typical human, such as the HAL computer in Arthur C. Clarke and Stanley Kubrick's film *2001: A Space Odyssey*, it has produced programs that perform some apparently intelligent tasks, often at a much greater level of skill and reliability than humans. More than this, AI has provided a powerful and defining image of what computer technology might someday be capable of achieving.

PETER M. ASARO

Further Reading

Anderson, J.A., and Rosenfeld E., Eds. *Talking Nets: An Oral History of Neural Networks*. MIT Press, Cambridge, MA, 1998.

Barr, A., Feigenbaum, E.A. and Cohen, P.R., Eds. *The Handbook of Artificial Intelligence,* vols. 1–4. Heuris Tech Press and Kaufmann, Stanford, (1981–89).

Crevier, D. *AI: The Tumultuous History of the Search for Artificial Intelligence*. Basic Books, New York, 1993.

Dreyfus, H. *What Computers Can't Do: A Critique of Artificial Reason*. Harper & Row, New York, 1979.

Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Englewood Cliffs, NJ, 1998.

Jurafsky, D. *et al. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ, 2000.

McCarthy, J. Programs with common sense. *Proceedings of a Symposium on Mechanical Thought Processes,* vol. 1. London: Her Majesty's Stationery Office, London, 1958, pp. 77–84. Reprinted in Minsky, M.L., Ed. *Semantic Information Processing*. MIT Press, Cambridge, MA, pp. 403–418, 1968.

McCorduck, P. *Machines Who Think*. W. Freeman, San Francisco, 1979.

Mitchell, T. *Machine Learning*. McGraw-Hill, New York, 1997.

Newborn, M. and Newborn, M. *Kasparov Versus Deep Blue: Computer Chess Comes of Age*. 1996.

Russell, S.J. and Norvig P. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ, 1995.
Shapiro, S.C., Ed. *Encyclopedia of Artificial Intelligence*, 2nd edn. Wiley, New York, 1992.
Stork, D.G., Ed. *HAL's Legacy: 2001's Computer as Dream and Reality*. MIT Press, Cambridge, MA, 1997.
Webber, B.L. and Nilsson, N.J., Eds. *Readings in Artificial Intelligence*. Morgan Kaufmann, San Mateo, CA, 1981.

Audio Recording, Compact Disk

The compact disk (CD) brought digital recording and playback technology within the reach of the average household. It offered a quality of sound reproduction good enough to satisfy the audiophile, unprecedented recording length within a compact format, and a system that did not damage sound quality with use. These qualities made it far superior to the micro-grooved vinyl disk for the reproduction of sound, and it rendered this long-lived technology obsolete within a decade of its introduction. The technology employed in this product has been applied to a variety of uses, from data storage and retrieval systems to an integrated audio and video (play only) disk for the home user.

Methods of saving and reproducing data have been a major area of research for engineers and scientists from the nineteenth century onwards. Many of the technologies developed have found applications in home audio—from Edison's revolving wax cylinders in the 1880s to magnetic recording tape in the 1940s. At every step, the accuracy of reproduction, the range of sound frequencies captured, and the durability of the recording medium have been progressively improved to reach its zenith with the compact disk.

In the 1960s and 1970s technologies from several different areas of this research were combined into a system of digital encoding and retrieval using optical readers to retrieve the data and computer sampling to turn analog sound into digital code. Much of the research was aimed at video recording, but it was soon applied to audio. Many scientists and business executives involved in this corporate research were audiophiles who wanted a superior method of saving and reproducing sound. The revolving vinyl disk and the system of stamping copies of master recordings provided important paradigms in the research effort. The compact disk was first introduced as a commercial product in 1982 by Sony Corporation of Japan and Philips electronics of the Netherlands.

The research that led to the CD was carried out in the U.S., Japan and Europe. James T. Russell, a physicist at the Battelle Institute, invented a system of using light to read binary code in the 1960s. By 1970 he had a digital-to-optical recording system using tiny bits of light and dark—each about 1 micron in diameter—embedded on a photosensitive platter. A laser read the tiny patterns, and a computer converted the binary code into an electrical signal that could be converted into audio or video streams. I.S. Reed and G. Solomon published a multiple error correction code in 1960 that would be employed in the encoding of data on CDs to detect and correct errors.

In Japan the NHK Technical Research Institute exhibited a pulse code modulation (PCM) digital audio recorder in 1967 that sampled sound and saved the binary data to videotape. Two years later Sony introduced a PCM recorder, and in 1978 it offered a digital audio processor and editor, the PCM-1600, to record companies and radio stations, which used them to make master recordings.

Credit for the idea of the compact disk is usually given to Klaas Compaan, a physicist working for the Philips Company. In 1969 he realized that an RCA system of stamping copies of holograms could be used to reproduce disks holding video images. With his colleague Piet Kramer he devised a glass disk on which they recorded video signals, along with a track of dimples to record the analog sound signal, that were read with a laser beam. They then moved to recording a digital code on the disk and used a digital-to-analog converter to reproduce sound from the encoded binary stream.

In the 1970s Philips, Sony, and several other companies introduced digital systems to save video and audio. In 1978 35 manufacturers of digital recorders met in Tokyo. Philips took the lead in establishing standards for the format of the audio system: the diameter of the disk was finally set at 120 millimeters; the sampling rate was to be 44.1 kHz; and a 16-bit standard was adopted for the encoding of the audio signal. This enabled 74 minutes of sound to be recorded on the disk. Manufacturers agreed to run the data track from the inside to the outside of the disk and use a polycarbonate material (developed by Polygram, a subsidiary of Philips) for the disk substrate.

Philips and Sony collaborated in the development of prototypes, and in 1980 they proposed a set of standards whose worldwide adoption was an important factor in the success of the compact disk. In 1983 CDs were first introduced in the U.S., and 800,000 disks were sold; in 1986 53,000,000 were sold. By 1990 an estimated 1 billion CDs were sold globally. Within a few years after the

introduction of the compact disk player, smaller units were available as car stereos and personal audio systems.

Several other products were quickly developed from the technologies used in compact disks. Compact video disks ranging from 120–300 mm in diameter were introduced in the 1980s. CD-ROM (Read Only Memory) units were developed as high-capacity data storage systems for computers. In the 1990s CD-I (Interactive) technology was introduced to merge interactive combinations of sound, pictures, computer texts, and graphics in one format. A combined audio and video 120 mm disk system (Digital Versatile Disk) was launched in 1996, which went on to successfully challenge the VHS video tape cassette recorder.

The compact disk quickly proved to be far superior to its vinyl distant cousin. Not only did it sound better and hold much more music, it was also free of the scratches and background noise that had become an accepted part of sound recording and reproduction. Yet it failed to dislodge the compact magnetic tape cassette from its preeminent position in home audio because it was a play-only format. The union of Sony and Philips was broken in 1981 and the two companies went their separate ways to develop compact disk technology and devise a suitable recorder. Several different systems were introduced in the 1990s ranging from Sony's Mini Disk recorder to the recordable CD-R used in home computers. The popularity of downloading MP3 files from the Internet and burning the sound onto recordable CD-Rs has made this format the most likely successor to the cassette tape.

See also **Audio Recording, Disks; Computer Memory, Personal Computer; Lasers, Applications; Personal Stereo**

ANDRE MILLARD

Further Reading

Everest, F.A., and Streicher, R. *New Stereo Sound Book*, Audio Engineer, New York, 1998.

Guterl, F. Compact Disc. *IEEE Spectrum*, 25 102–108, 1988.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, New York, 1995.

Pohlmann, K. *The Compact Disc Handbook*. A-R Editions, Madison, 1992.

Reed, I.S., and Solomon, G. Polynomial Codes Over Certain Finite Fields. *J. Soc. Appl. Math.*, 8, 300–304, 1960.

Read, O., and Welch, W.L., *From Tin Foil to Stereo: Evolution of the Phonograph*. SAMS, Indianapolis, 1976.

Audio Recording, Electronic Methods

The mechanical method of recording sound was invented by Thomas A. Edison in 1877. With the help of mass production of recorded copies for entertainment, cylinder phonographs, and then disk phonographs developed into a major industry during the first quarter of the twentieth century. However, electronic amplification was not available. Modern research has shown a few examples of *pneumatic* amplification, and there were also experiments using electronic approaches. The first published record known to have been made with the help of an electronic amplifier was Guest and Merriman's recordings at the Burial Service of the Unknown Warrior in Westminster Abbey on November 11, 1920 (specifically for the Abbey Restoration Fund and not a commercial issue).

Several components were needed by experimenters: a microphone, an electronic amplifier, a loudspeaker system to hear what was being recorded and to avoid overloading, and an electromechanical device which would faithfully transmute the amplified signal into mechanical vibrations for cutting a groove in wax with negligible added background noise. Cutting records and the means for pressing sturdy disk records were commercial trade secrets at that time.

The vital breakthroughs occurred at the American Telephone and Telegraph Company (AT&T) after World War I, when their research section Bell Laboratories began studies for improving transcontinental telephone communication. E.C. Wente developed a microphone in 1916, and in 1924 Henry C. Harrison developed an elaborate theory of "matched impedance" for sending speech long distances without losses. He realized that the same principles could be used for the faithful reproduction and recording of mechanically recorded sound. His team translated these electrical studies using "analogies." For example, an electrical capacitance might be regarded as analogous to a mechanical spring, or the springiness of air in a confined space. Based on these analogies, the team designed a complete system, including microphone, amplifier, loudspeaker, and cutter; *and* an improved mechanical reproducer for judging the results.

The ideas were marketed by AT&T's manufacturing and licensing company Western Electric. As far as can be established, the Western Electric recording system was first used commercially in New York on February 25, 1925. A few earlier examples are now known, and some have been published.

The Western Electric amplifier was originally developed for public-address purposes, and it used electronic valves, or vacuum tubes. Several other recording systems with electronic amplification immediately appeared, demonstrating that parallel research had occurred; but a relatively small amount of electronic amplification can often compensate for classical thermodynamic inefficiencies. Among the new systems were the inventions of P.G.A.H. Voigt for the British Edison Bell Company and Captain H. Round for the British Marconi Company, both of whom developed alternative microphones.

A microphone changes sound into alternating electricity and should introduce little distortion to the frequencies picked up. It should have the same frequency response in all directions, and its linearity to acoustic waveforms should be uniform. However, by the end of the twentieth century, electronic amplification had not enabled extremely faint sounds (detectable by a healthy human ear) to be recorded. There was always added random noise arising from acoustic causes, electronic causes, and inefficiencies in acoustical–mechanical transducers. Microphone users had to select an instrument suited to the proposed application.

Although the earliest optical film sound experiments probably did not use electronic amplification, both Western Electric and the merged RCA-Victor Company developed independent methods of recording sound on optical film with the help of electronics. Here, the two methods had to be "compatible" so that any film could be shown in any theater, and this situation remained true with only minor exceptions until the 1980s.

Film studios provided two new electronic techniques widely understood today but much less so before the 1960s. "Automatic volume limiting" protected the fragile "light valves," with the side effect of improving speech intelligibility for cinema reproduction. The difficulties of recording foreign languages led to what is now called "multitrack recording," in which two or more sounds recorded at different times could be kept in synchronism and modified as necessary. Hollywood evolved the principle of three synchronous soundtracks for music, sound effects, and dialog, to facilitate the adaptation of films for foreign markets.

The second technique was "negative feedback." This principle enabled a high degree of amplification to be traded for other purposes by feeding some of the output (reversed in phase) back to a previous stage. In this process, nonlinearity and deviations in frequency response may be reduced. Mechanical transducers could be built with "motional feedback" with the same results. In sound recording, motional feedback was first used in disk equipment for the National Association of Broadcasters in America. This was a different system from that used in commercial recording, and because of the outbreak of World War II, the commercial implementation of this principle was delayed until 1949. Cutting stereophonic disks would have been impossible, however, without motional feedback.

Classical analog information theory stated that frequency range could be traded against the need for amplification. At the start of the 1930s, both films and mechanical disk records were restricted to an upper frequency range of about 5 kHz. This range was gradually extended by trading amplification for electromechanical efficiency. It is generally accepted that the full audio frequency range was first achieved by the English Decca Company during 1944 (as a spinoff from war research). An upper limit of 14 kHz was obtained for microphones, amplifiers, and disk pickups.

During World War II, German engineers rediscovered a patented American invention for reducing harmonic distortion on magnetic media due to hysteresis. The principle of ultrasonic alternating current bias greatly improved the linearity and reduced the background noise of magnetic tape, causing much debate about precisely how the principle worked, which still has not received a complete explanation. In analog sound recording, a useful element of feedback happens with this process. If a small particle of dirt should get between the tape and the recording head, the bass goes down and the treble rises. When the electronics are correctly set up, these effects cancel each other; and as a result, analog magnetic tape became the preferred mastering format for professionals. This format required even more amplification; but by the mid-1950s, magnetic tape was much simpler and cheaper than mechanical or optical recording, so it remained the favorite technology for analog recording until the end of the century. Digital audio recording using magnetic tape and optical methods began to overtake it after about 1980, but no fully digital microphone was developed.

See also **Audio Recording, Compact Disk; Audio Recording, Disks; Audio Recording, Stereo and Surround Sound; Audio Recording, Tape; Film and Cinema, Early Sound Films; Film and Cinema, High Fidelity to Surround Sound; Valves/Vacuum Tubes.**

PETER COPELAND

Further Reading

Borwick, J., Ed. *Sound Recording Practice*, 4th edn. Association of Professional Recording Studios, Oxford University Press, Oxford, 1994.

Copeland, P. *Sound Recordings*. British Library, London, 1991.

Langford-Smith, F. *Radio Designer's Handbook*, 4th edn. RCA, U.S., 1953. Chapters 12 to 21 cover all aspects of "Audio Frequencies."

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, Cambridge, 1995

Morton, D. Off The Record: The Technology and Culture of Sound Recording in America. Rutgers University Press, New Brunswick, 2000.

Audio Recording, Mechanical

Mechanical recording and reproduction of sound on disk followed ten years after Thomas Edison's invention of these techniques on cylinder. In 1887, Emile Berliner filed for and received his patent on the gramophone recording and reproduction system. Shortly after 1900, technical developments and commercial agreements in the U. S. established disk records as an independent medium. By 1910 they dominated global markets for recorded sound. Mechanical recording remained the technique of choice for the storage of sound for the next 40 to 50 years, and disk "records" continued to be a significant medium for consumers of music and other sounds into the 1980s.

Disk recording requires that sound waves be converted into mechanical force that makes a stylus engrave or otherwise reproduce the waves as a spiral groove on a spinning disk composed of some pliant material. To play back the sound wave on the same disk or on a copy stamped from a matrix, a stylus tracks the groove representing the wave. This energy is then converted into mechanical force that drives one or more loudspeakers. The development of this technology during the twentieth century is one of incremental improvements punctuated by revolutionary changes in technique and materials.

Recording changed very little in form from 1901, when Eldridge Johnson's synthesis of earlier inventors' work with his own insights became the Victor Talking Machine Company, to 1925. Performers directed their voices or instruments at a horn that channeled sound waves to a diaphragm. The membrane was linked to a stylus driven by a feed screw along the radius of a metallic wax disk spinning between 74 and 82 revolutions per minute (rpm). Miniature flyball governors controlled the speed of the turntable so that the record was rotated at a constant speed. The stylus cut the wax vertically ("hill and dale") or laterally.

Once recorded, the disk was electroplated, creating a master from which records or secondary stampers could be made. Victor and its partner companies made records from mineral powder, shellac, and carbon black; Columbia Records either copied Victor's technique or used pure shellac laminated over a paper core. Edison had also moved from cylinders to records, and the "Diamond Disks" made from 1912 to 1929 used a wood flour base coated with a varnish called condensite. Consumers played records on spring-wound gramophones or talking machines where the process for recording was reversed through steel stylus, diaphragm, tone arm, and horn.

Improvements in diaphragms, styluses, room acoustics, and horn placement enabled increases in the signal-to-noise ratio and frequency range while reducing distortion in the reproduced sound. By the late 1910s, companies had standardized record diameters at 10 and 12 inches (254 and 305 mm), playing three and five minutes respectively, and progressed from recording individual performers of selected instruments to symphony orchestras.

The revolution of electronic recording with microphones and amplifiers took place after World War I. Joseph Maxfield and Henry Harrison of Bell Telephone Laboratories applied scientific techniques of research and development to extend the undistorted frequency range of recording from 220–4,000 Hz to 100–6,000 Hz, and to raise the signal–noise ratio to 32 dB. To meet competition from wireless broadcasting, record companies began licensing the Bell system in 1925.

Electronic recording stimulated systems improvements and standard speeds in reproduction. For the consumer and home market, the "78" record (78.26 rpm) was standard, and the broadcast and film industries used 33.33 rpm. Electronic amplification reduced the need for abrasive fillers that resisted the stylus pressure of mechanical reproduction, while the higher frequencies aggravated the sound of needle "scratch" generated at the interface of stylus and the record groove.

Despite the global economic depression, the 1930s were a fertile period for recording innovations. Alan D. Blumlein of Electric & Musical Industries and Arthur C. Keller of Bell Laboratories invented single-groove stereophonic recording systems, for which their corporations could not find markets. Record companies began recording on metal disks coated with man-made lacquers and stamping disks of man-made plastics.

Researchers learned more about acoustics, recording and plating techniques, and the science of the stylus–groove interface. In the latter field, Frederic V. Hunt and John Pierce of Harvard University developed a stylus and pickup ten times lighter than contemporary devices.

Hunt and Pierce also anticipated the advantages of the narrower microgrooves that RCA Victor engineers began applying in 1939 to their vinyl chloride–vinyl acetate copolymer, 7 inch, 45 rpm, record system. World War II and rising sales deferred replacement of the 78 until 1949, a year after Columbia Records introduced vinyl compound, 10- and 12-inch, 33.33 rpm long-playing records (LPs) using RCA's microgrooves. The jukebox industry quickly adopted the 45, which survived to the end of the century in the pop singles market. The new records gave less distorted sound reproduction up to 10,000 Hz with a 60 dB signal–noise ratio.

At the time, all music was monophonically recorded and reproduced. The postwar growth of magnetic recording techniques enabled multitrack recording, which led to the international standard for stereophonic disk reproduction in 1958. During this time, engineers refined the art of microphone placement and recording techniques to produce what has become known as the "Golden Age" of recording.

Efforts to put four channels on disk in the early 1970s failed because of the lack of a standard and the rise of magnetic-tape cassettes. Cassettes surpassed LP sales in the United States in the late 1970s, and this was followed by the introduction of the compact disk (CD) ten years later. Major labels abandoned vinyl records in 1990, although some production continued in the U.K. and on independent record labels.

The main refuge for disk records became audiophile and reissue pressings and discotheques, where disk jockeys had been playing extended versions of popular songs on 12-inch records since the early 1970s. The playing time of 5 to 12 minutes permitted increased amplitude of the groove and a concomitant rise in dynamic range. Ironically, in the middle of the digital revolution of CDs, the cult success of vinyl disks forced the major companies to continue pressing LPs and 45s for top performers through the end of the century.

See also **Audio Recording, Compact Disk; Audio Recording, Electronic Methods; Audio Recording, Stereo and Surround Sound; Audio Recording, Tape; Personal Stereo**

ALEXANDER B. MAGOUN

Further Reading

Gelatt, R. *The Fabulous Phonograph, 1877–1977*. Macmillan, New York, 1977. While somewhat dated, still the best single narrative.

Gronow, P. and Ilpo, S., translated by Moseley, C. *An International History of the Recording Industry*. Cassell, London, 1998.A broader survey, focusing on performers.

The phonograph and sound recording after one-hundred years. *J. Aud. Eng. Soc.*, 25, 10/11, October/November 1977. With survey articles on aspects of the phonograph system edited by the former chief engineer for RCA Records.

Guy M. and Andrews F., Eds. *Encyclopedia of Recorded Sound in the United States*. Garland, New York, 1993. An ambitious one-volume compendium that covers the art and science of the subject.

Read, O. and Welch, W.L. *From Tin Foil to Stereo: Evolution of the Phonograph,* 2nd edn. SAMS, Indianapolis, 1976. Considered the best history of the phonograph.

Audio Recording, Stereophonic and Surround Sound

It is perhaps difficult to understand the problems experienced by early listeners when they first heard "music coming out of a hole." It was an effect that annoyed some people much more than others, although few had even encountered it before the twentieth century. Together with the difficulties of low amplification and high background noise, this effect proved significant and was a further problem in sound recording and reproduction.

In the late twentieth century, the words "stereophonic" and "stereo" refer to a sound along a horizontal line, usually between two loudspeakers along one wall of a room. Good stereo can simulate a live performance, but the expression is restricted to loudspeaker listening. Listening on headphones is quite different, because our hearing has evolved to take advantage of two different phenomena. For natural sounds (and for loudspeakers), our heads form a "sound shadow," so high-pitched sounds from the left are picked up by the left ear very much louder than the right ear. For low-pitched sounds, the shadow-effect is practically nonexistent. What little directionality we possess depends upon differences in the *time* the sound arrives at the two ears. If recordings were made with microphones the same distance apart as our ears (about 17 centimeters), and something the same general shape and size as a human head were put between the microphones, we might have much more faithful stereo reproduction than any loudspeakers could give. This is called "dummy head" or "binaural" stereo. We learn directional hearing

in the first year of life, correlating the senses of sight and stereo sound with feelings of balance and the sensations our muscles give us as they move. With headphone listening, the acoustic fidelity is certainly better; but when the listener moves his head, the earpieces move too and the sound remains the same, yet his other senses indicate the exact opposite.

The first publicized experiment into spatial sound reproduction was performed by Clement Ader in Paris in 1881. He placed a row of microphones along the footlights of the Paris Opera House and sent signals by telephone lines to earpieces at an electrical exhibition 3 kilometers away. A spatial effect was reproduced. American workers were first to achieve stereo sound recording. Closed-circuit experiments established that just one pair of microphones and loudspeakers would reduce the disadvantage of "sound coming out of a hole;" but attempts to reproduce physical movement showed that if the microphones were more than a few feet apart, a new "hole in the middle" might appear. A third central microphone-and-loudspeaker arrangement reduced this "hole in the middle," demonstrating that three were much better in practice.

In 1932 a significant experiment was performed using landline connections in Philadelphia, which seemed to show that for orchestral music, three channels would be sufficient. While significant, it was not a definitive experiment because there were apparently no comparisons with more channels and the vertical element was not featured. Nevertheless this established the benchmark against which future systems would be compared. This experiment led to the first stereo recordings in 1932 of the Philadelphia Orchestra conducted by Leopold Stokowski, recorded on a two-track disk at Bell Laboratories in New Jersey. Because there was only one experimental stereo disk cutter, continuous performances could not be recorded. In 1979 Bell Labs reissued the surviving material on two charity stereo long-playing records (LPs) with the missing sections in monophonic sound, clearly showing the advantage of stereo.

On the other side of the Atlantic, work started from different basic principles. It was noticed that when mono sound was reproduced on two identical loudspeakers equidistant from the listener, the sound appeared to come from a point midway between the two speakers. By dividing the same mono sound into suitable proportions and sending it to two speakers, it could appear to come from anywhere between the two. This was the claim in Alan Blumlein's British patent in 1931. Blumlein was an employee of the record company EMI, but he considered the principal application to be soundtracks for films, unlike in the U.S. where the initial work was for sound alone. Blumlein's work showed how suitable signals might be captured by two bidirectional microphones at the same point in space. It also showed how the correct balance for monophonic listeners would result from mixing the left and right in equal proportions, so downward compatibility would occur. It also discussed how stereo sounds might be recorded on disk or film, including some remarkable predictions of what occurred decades later. The important point was that Blumlein showed that only two discrete channels could be sufficient.

Blumlein's work also mentioned a device to "steer" a mono sound between two loudspeakers, now called a "panpot." In 1934 the EMI Amateur Dramatic Society made a short stereo film called *Move The Orchestra*, in which a single restaurant scene achieved exactly that.

The public heard stereophonic sound for the first time in 1939 in the Walt Disney film *Fantasia*. In the "Fantasound" system, three discrete channels were printed onto a separate optical film that held the three soundtracks side-by-side, together with a fourth track that controlled the volumes of the other three, giving wider dynamic range. During the last musical piece (*Ave Maria*), loudspeakers at the back of the auditorium could be switched on for the choral effect. This was not only the first time the public had heard stereo, but it was the first time they encountered "surround sound." The outbreak of World War II brought all these advances to a halt.

After LPs became established in the 1950s, techniques were researched to get two channels of sound into one groove. For commercial master sound recording in the U.S., the three traditional discrete channels were initially recorded on $\frac{1}{2}$-inch magnetic tape. In Europe, EMI issued the first British "stereosonic" $\frac{1}{4}$-inch magnetic tapes in 1955 using Blumlein's principles. They were intended for loudspeaker listening; Blumlein's "coincident microphone" technique did not capture the time information needed for headphones. The standard form of stereo disk that carried two sounds in one v-shaped groove was launched in 1958. A single cutter vibrating in two directions mutually at right angles imparted two sounds onto the two groove walls, which gave compatibility between mono and stereo disk reproduction.

Although human eyes see an almost circular field of vision, most practical action takes place on a narrower plane. Widescreen filmmakers followed

American thinking: a different microphone for each loudspeaker behind the screen. But dialog was found to give problems since it was upset both by performance considerations and the acoustic properties of studios. Subsequent stereo films used sound picked up on mono microphones and "panned" into place. Ultimately, mono became standard for dialog, with only music and effects in stereo.

During the first half of the 1970s, the high fidelity, or hi-fi, market stimulated the idea of "quadraphonic" sound reproduction. The idea was a commendable one, to reproduce the entire acoustic environment of a concert hall rather than just the stage. The concept failed for several reasons, but it can be argued that the sophisticated engineering research needed for quadraphonic LPs helped prolong the life of mechanical disk records until the end of the century.

Until digital methods came to the rescue, the shortage of analog channels continued to hamper surround sound. Dolby Laboratories had invented methods of improving the perceived signal-to-noise ratio of analog recording media, and they also found ways to encode more channels of audio with extra information for cinema audiences, while retaining downward compatibility with stereo and mono. The first "Dolby Surround" films date from 1976.

When video recording became suitable for domestic use, the same principles became available to domestic listeners. At the end of the twentieth century, there were several ways of encoding surround sound. The dominant one used 5.5 channels (the "half" being low-pitched sounds such as aircraft, guns, and bombs normally requiring a large loudspeaker, which could be placed behind the viewer). It seems certain that further developments will occur in the twenty-first century. The *vertical* element of surround sound, for example, is not yet featured, although it is an increasingly important part of our environment.

See also **Audio Recording, Electronic Methods; Film and Cinema, High Fidelity to Surround Sound; Loudspeakers and Earphones; Personal Stereo; Television Recording, Tape**

PETER COPELAND

Further Reading

Copeland, P. *Sound Recordings*. British Library, London, 1991.

Hadden, H.B. *Practical Stereophony*. Iliffe Books, London, 1964.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, Cambridge, 1995.

Rumsey, F. *Spatial Audio*, Focal Press, 2001.

Audio Recording, Tape

Tape recorders were the dominant technology for sound recording in the second half of the twentieth century. By combining low cost, high fidelity, and flexibility in editing, they came to dominate both consumer and professional recording during this period. In addition to sound, tape recorders were used for video and data recording, which are discussed in separate entries.

Tape recorders are characterized by the use of a recording medium with a flat cross section. The tape is moved rapidly past a recording head, which generates a magnetic field in response to an electrical input, normally from a microphone. The magnetic field leaves the tape permanently magnetized. The field on the tape is then played back by moving the tape past a reproducing head, which senses the field and produces a signal that is amplified to recreate the original sound.

Valdemar Poulsen, inventor of the first wire recorder in 1898, also experimented with tape recorders around 1900. His machines used a solid metal tape, which provided superior performance since the recording medium always had the same geometrical relationship to the reproducing head as the recording head (wire could twist about its axis, resulting in variable sound quality). However, solid metal tape was much more expensive than wire, making tape recorders too costly for most applications.

As a result, solid metal tape recorders were built only for highly specialized applications during the 1920s and 1930s. Lorenz in Germany and Marconi in the U.K. built radio broadcast recorders using solid metal tape, and AT&T in the U.S. built telephone announcing systems, but their high cost meant that only a handful of each type of machine entered service. The key to more widespread use of tape recorders was the development of particle-coated tape by the Austrian inventor Fritz Pfleumer in 1928. By coating paper tape with small magnetic particles, he dramatically reduced the cost of magnetic tape. Coated tape was also physically much more flexible than solid metal tape, and this made the design of transport mechanisms for the tape much simpler.

Pfleumer sold his invention to the German firm AEG, which then partnered with the BASF division of I. G. Farben to develop a tape recorder. AEG developed and manufactured the machine

itself, while BASF developed and manufactured the tape. BASF soon replaced paper with plastic as the backing material, further lowering costs. The AEG machine was first marketed as the Magnetophone in 1935. The firm developed a variety of models for civilian, government, and military use over the next ten years. The Nazi government made extensive use of the Magnetophone for both security and radio broadcast purposes, and the demand for higher quality reproduction drove the development of a number of innovations. The most significant of these were the invention and widespread use of the ring head, which produced more intense and defined magnetic fields on the tape, and AC-bias, an electronic noise reduction technique. AC-bias was the addition of a very high frequency tone to the input signal during recording, which dramatically reduced noise and made tape recording suitable for music as well as voice recording.

The developments at AEG and BASF were copied in the laboratory in the U.S. and the Soviet Union during World War II, but it was not until the general dissemination of knowledge about the Magnetophone after Germany was occupied by Allied forces that there was widespread interest in tape recording. The Brush Development Company in the U.S., under the technical leadership of Dr. Semi Joseph Begun, produced the first American tape recorder, the Soundmirror, in 1947. Brush was soon followed by a number of other companies, most notably Ampex, which produced a high-quality recorder suitable for radio broadcast use. Using a combination of German technology and American wartime research, these and other firms in the U.S., Europe, and Japan soon offered a variety of products whose performance matched or exceeded competing recording technologies.

By the early 1950s, tape recorders were the dominant method for making professional sound recordings. They had rapidly replaced other forms of recording in the motion picture industry and were increasingly used in radio broadcasting and the music recording industry. The success of tape recording was due not only to lower cost but also to the ease of editing. Unlike phonograph recordings, magnetic tape could easily be cut and pasted together, much like motion picture film, to make a high quality product. Such editing had been possible with sound-on-film systems, but magnetic recordings could be edited immediately rather than waiting for film to be developed.

The popularity of tape recording for professional use increased further in the 1960s with the development and widespread use of multitrack tape recorders. In addition to the basic ability to separate vocal and instrumental tracks, these recorders provided much easier sound editing in the studio and also allowed artists to experiment with new forms of musical expression. Analog multitrack recorders began to be replaced in the mid-1970s by digital multitrack recorders for studio use. By the end of the twentieth century, most professional recording was digital and increasingly recorded on hard disks or other computer-related formats rather than on tape recorders.

The use of tape recorders by consumers lagged behind that of the professional world. Reel-to-reel tape recorders, the primary consumer versions of the 1950s, were relatively inconvenient to use compared with phonograph record players and other consumer electronics and were usually limited to fixed installations in the home. A variety of cassette formats were developed to address this problem, but it was not until the introduction of the Phillips Compact Cassette (so common by the end of the twentieth century that it was simply referred to as "the cassette" as if it were the only one) and the 8-Track Cartridge in the 1960s that cassettes were widely adopted.

The 8-Track was the first successful high fidelity cassette system, and it was installed in many vehicles in the 1960s and 1970s. Because of the greater physical size and complexity, it eventually lost out to the Phillips cassette. The Phillips format was originally intended for voice recording only, but the development of new magnetic particle formulations in the 1970s, along with the introduction of Dolby® noise reduction, improved sound quality of the Phillips cassette to the point that it matched the performance of the 8-Track and the LP phonograph record and so gained consumer acceptance. At the end of the twentieth century, tape recorders and especially the Phillips cassette were still the dominant form of consumer recording technology, but they were increasingly being supplemented by newer digital recording formats primarily derived from the computer industry.

See also **Audio Recording, Wire; Computer Memory; Television Recording, Tape**

MARK H. CLARK

Further Reading

Daniel, E.D. and Mee, C.D. *Magnetic Storage Technology*, 2nd edn. McGraw-Hill, New York, 1996.

Daniel, E.D., Mee, C.D. and Clark, M.H. *Magnetic Recording: The First 100 Years*. IEEE Press, 1999.
Morton D. *America on Record*. Rutgers University Press, New Brunswick, 2000.

Audio Recording, Wire

The recording of sound using the principle of remnant magnetism (the residual magnetic fields present in some materials after they are exposed to magnetic fields of sufficient strength) was first proposed in 1878 by the American inventor Oberlin Smith. His ideas, later recorded in a patent caveat, are the intellectual starting point for the development of all modern magnetic recording systems.

Wire recorders are characterized by the use of a solid recording medium with a small circular cross section. The wire is moved rapidly past a recording head, which generates a magnetic field in response to an electrical input, normally from a microphone. The magnetic field leaves the wire permanently magnetized. The field on the wire is then played by moving the wire past a reproducing head, which senses the field and produces a recorded signal that is amplified and played back to recreate the original sound.

The primary advantage of the wire-recording medium is its low cost as compared with solid metal tape; machinery used in piano wire manufacture was easily used to produce wire for magnetic recording. Wire has many disadvantages however. In contrast to tape, which always presents the same orientation to the reproducing head, wire can twist in the transport mechanism, leading to variations in reproduction volume. As a result, wire recorders can record only a single track, in contrast to the multiple track possibilities of tape recorders. Wire also is harder to splice than tape, and it easily becomes snarled if it escapes from the reels on which it is wound. Finally, steel with ideal magnetic characteristics is very stiff and difficult to pull through transport mechanisms. This problem can be dealt with by using coated wire (developed in the early 1940s), although the higher cost reduces wire's economic advantages.

The primary application of wire recorders is sound recording. Recordings are analog recordings; that is, the magnetic pattern produced on the wire is an analog copy of the original sound. Wire recorders were also occasionally used for the digital recording of telegraph and radio telegraph signals early in the twentieth century.

The first successful magnetic recorder of any kind was a wire recorder constructed by the Danish inventor Valdemar Poulsen in 1898. Poulsen had been attempting to develop a telephone answering machine, and so he named his invention the Telegraphone (roughly meaning "distant sound recorder)." With the assistance of Danish investors led by the entrepreneur Lemvig Fog, Poulsen began to offer commercial machines in 1903. By that time, Poulsen had abandoned magnetic recording for research on radio transmission, and he and his collaborators turned further development of the telegraphone over to the American Telegraphone Company in the U.S.

After delays caused by financial and technical problems, wire recorders were offered in small numbers to the public by American Telegraphone during the 1910s and 1920s. Total sales amounted to fewer than 300 machines due to high prices, poor sound reproduction quality, and worse quality control. American Telegraphone entered bankruptcy proceedings in 1919 and ceased manufacturing in 1928, though the firm was not formally out of business until 1941.

During the late 1920s and 1930s, the primary manufacturer of wire recorders was the Echophone Company in Germany, later sold to Lorenz. The Textophone (later renamed the Dailygraph) was similar in design to the machine produced by the American Telegraphone Company, but incorporated an electronic tube amplifier to improve sound quality. The quality of the machines was higher than American Telegraphone products, but prices were just as high. As a result, the device was a niche product, selling primarily to wealthy businessmen interested in cutting-edge technology until 1933. With the rise of the Nazi Party to power after that date, Lorenz sold an increasing number of machines to the Gestapo and other state security agencies in Germany for wiretaps and the recording of interrogations. However, by 1939 competition from tape recorders produced by AEG led to the end of wire recorder production in Germany.

Although several organizations in the Soviet Union and firms in the U.S., most notably AT&T through its Bell Telephone Laboratories subsidiary, experimented with magnetic recording during the 1930s, production of wire recorders for consumer use did not take place until World War II in either country. Thus, at the same time wire recorder production was ending in Germany, it was increasing dramatically in the USSR and the U.S. However the military provided the major market for wire recorders in both countries. The primary application was in situations where vibration rendered other forms of recording impossible. Wire recorders were used primarily by reconnais-

sance pilots to record their comments while in flight and by journalists reporting from frontline locations.

Most military recorders used cassette designs to eliminate problems with threading the fine wire through a recording head. Cassettes also minimized the chance of wire snarls because of wire coming off the reels. Improved recording heads and the development of coated wire (an outer magnetic layer electroplated on a more flexible nonmagnetic inner core) meant that the sound quality of wartime wire recorders was much better than before. However, the most notable improvement in sound quality resulted from the introduction of alternating current (AC) bias, an electronic noise reduction technique originally developed for tape recorders.

The Armor Research Foundation was the leading wartime developer of wire recorders. Armor patented much of its work and then licensed its patents to dozens of firms in the U.S. and abroad in the late 1940s. These firms produced essentially all the wire recorders aimed at the consumer market in the U. S. and Europe after World War II. Annual production of consumer wire recorders peaked at several million around 1950.

In contrast to the high quality of military recorders built during World War II, postwar consumer wire recorders were largely inexpensive units built as cheaply as possible. In particular, very few of them incorporated cassette mechanisms. As a result, snarls and other malfunctions were a constant problem. Consumers put up with these difficulties since wire recorders sold for considerably less than tape recorders in the immediate postwar period. Wire recorders vanished from the consumer market by the mid-1950s as the price of tape recorders using coated plastic tape dropped. Although they continued to be used in special applications through the 1960s, wire recorders have essentially disappeared with the development of improved tape materials.

See also **Alloys, Magnetic; Audio Recording, Tape**

MARK H. CLARK

Further Reading

Daniel, E.D. and Mee, C.D. *Magnetic Storage Technology*, 2nd edn. McGraw-Hill, New York, 1996.

Morton D. *America on Record.* Rutgers University Press, New Brunswick, 2000.

Daniel, E.D., Mee, C.D. and Clark, M.H. *Magnetic Recording: The First 100 Years*. IEEE Press, 1999.

Audio Systems

At the start of the twentieth century, mechanical recording and reproduction of sound by cylinder or disk phonographs was dominant (Edison's cylinder, Edison's Diamond Disk, Emile Berliner's gramophone, and the Victor Talking Machine Company's Victrola). Playback quality was limited by the lack of amplification methods to increase the volume of the recorded performance above the background noise of the mechanical record. Thus inexperienced performers were rarely used. Electronic recording methods with microphones and amplifiers were developed after World War II, and valve amplifiers and the transistor, which created more efficient signal amplification, caused a revolution in audio playback.

Throughout the twentieth century, turntables or decks for playing disk records were normally kept horizontal, so the downward pressure of the stylus would be constant across the disk's surface. Eldridge Johnson, founder of the Victor Talking Machine Company in 1901, devised a spring motor to rotate the turntable at an even speed in 1896, and battery motors were also used prior to the electrical drives developed in 1925. Although disk records were mastered on lathe-type mechanisms that moved the cutter in a straight line across the disk, records were usually played with a pivoted "pickup arm" that made a curved rather than straight track across the record, introducing distortion as well as wearing the record. Careful geometric design and motors that moved the rear end of the arm reduced tracking error and improved audio quality.

Early playback was acoustomechanical, as was sound recording. Following the development of electronic sound recording, the first fully electrical playback device, the Brunswick "Panatrope" in 1926, used magnetic playback pickups (cartridges) to convert the motion of the stylus in the groove to an electrical signal. Brunswick's player also incorporated an electronic amplifier using the new vacuum tubes, and the electrical signal could be played through a loudspeaker.

The earliest material for disk records was called "shellac." Actually a mixture of materials, shellac was an organic binding material for which no synthetic substitute was found. Before the first electronic version, sounds were reproduced and amplified mechanically. The acoustic energy was emitted by a "soundbox" and fed into a conical horn providing a suitable acoustical load. Ultimately, this energy came from the rotation of the disk itself. Therefore the disks had to resist

wear; and to this day, shellac has proven to be one of the longest lasting sound storage media. Shellac was used until about 1955 in the U.S., 1959 in the U.K, and even later in developing countries. The major drawback to shellac is that it is very prone to breakage.

During World War II the Allied forces made effective use of music for maintaining the morale of the troops, and music had to be distributed without risk of breakage. Shellac was gradually replaced with "vinyl" (polyvinyl chloride), which was flexible and had much less "grain," meaning lower background noise. Until then, heavy steel needles, or pickups, were used for playing disks, typically with downward pressures of 100–150 grams. When long-playing disks (LPs) were introduced after the war, the same vinyl was used. It was much more expensive than shellac, but the longer playing time offset this. Pressures from 10 to 20 grams became normal, with reduced stylus tip size. Vinyl disks also featured extended frequency range, and the extended range of good pickups also reduced wear. Also, because of vinyl's inherent flexibility, turntable designers were forced to introduce a pickup with a carefully crafted playback stylus made of jewel such as sapphire to resist wear. A number of mathematical studies showed how to optimize various dimensions of quality sound to achieve faithful reproduction, low distortion, and simultaneously low damage to disks. Vinyl records were still used by enthusiasts at the end of the twentieth century.

The faithful capturing of sound performed under "normal" conditions was led by radio broadcasting in the early 1920s. The technology of the electronic amplifier aided both the recording and the reproduction of sound. The first electronic amplifiers were based on thermionic valves, or vacuum tubes. Since vacuum tube technology was familiar to radio engineers, the principal developments in both radio and audio work tended to be described in books and magazines for the radio market. The demobilization of armed forces after World War II and expansion of the consumer market seems to have led directly to pure audio research and literature.

From 1950 onward, the pursuit of innovative, if not always faithful, sound reproduction was fed by specialist manufacturers, dealers, exhibitions, and advertising claiming "high fidelity," or hi-fi, playback. A disk deck tended to form the centerpiece of a practical domestic hi-fi system. Tape decks could also used for playing reels of magnetic tape, with cassette decks added for analog audiocassettes after 1970. A tuner, primarily for receiving analog frequency-modulated (FM) radio broadcasts, was a normal component.

The actual amplification took place in two or three separate units, depending on whether monophonic or stereophonic reproduction was intended. The preamplifier brought the various sources of audio to similar electronic voltages, allowed the bass and treble ranges to be controlled for individual preference and, in the case of the disk deck, applied equalization. This compensated for deliberate changes in frequency response to improve the capacity of a mechanical disk record, called a recording characteristic, internationally standardized in 1955.

In vacuum tube days, the main amplifier (or amplifiers for stereo) tended to be a heavy, delicate, and hot unit, so it was usually placed on the floor out of the way. It had enough output power to feed a separate loudspeaker enclosure, or two for stereo. Hi-fi loudspeakers might require between 5 and 50 watts apiece. Most 1950s designs were no more efficient than those at the end of the century, the electroacoustic efficiency being only about 1 to 3 percent. Quality transistorized amplifiers became dominant toward the end of the 1960s.

Each loudspeaker cabinet often contained several electrical-to-acoustical transducers covering different parts of the frequency range, with the audio being suitably divided by a "crossover unit." Compared with anything else in the high-fidelity system, loudspeakers had the most radical effects upon the reproduced sound quality. They were relatively cheap and simple for enthusiasts to assemble, and many did.

As the actual hardware became cheaper, with greater appeal to nonspecialists, the hi-fi craze became democratized. No longer was it confined to the wealthy or dedicated enthusiast. By the end of the twentieth century, everyone could enjoy high fidelity sound very comparable to leading technology of the 1950s.

See also **Audio Recording, Disks; Audio Recording, Electronic Methods; Audio Recording, Tape; Audio Recording, Wire; Loudspeakers and Earphones; Personal Stereo**

PETER COPELAND

Further Reading

Copeland, P. *Sound Recordings*. British Library, London, 1991

Crabbe, J. *Hi-Fi In The Home*, Blandford Press, 1968.

Day, T. *A Century of Recorded Music: Listening to Musical History*. Yale University Press, New Haven, CT, 2001.

Read, O. and Welch, W.L. *From Tin Foil to Stereo: Evolution of the Phonograph*. SAMS, Indianapolis, 1976.
Wilson, P. and Webb, G.W. *Modern Gramophones and Electrical Reproducers*. Cassell & Company, 1929.

Audio Systems, Personal Stereo, *see* **Personal Stereo**

Audiology, Hearing Aids

Treatment and testing for hearing disorders can be traced back as far as the first century BC. However, nothing prior to the sixteenth century indicates that any consistent use or application of hearing assistance devices existed.

The fundamental function of early hearing aids was to amplify sound. Before electricity, the only way to do that was to filter out other noise by directing the desired sound straight into the ear with some kind of tube or trumpet. Ear trumpets were first used by sailors and others who needed to communicate over long distances, and later on were adopted for use by the hearing impaired. Most early trumpets were custom made, and the real business of manufacturing and selling hearing aids only began around 1800. By the end of the nineteenth century a large variety of ear tubes and trumpets were available in many styles and designs, ranging from cheap devices made of tin or hard rubber to the more expensive ones constructed of more valuable materials. The more expensive models were often treated like jewellery but did not necessarily work better. Cupping the hand behind the ear makes sounds 5 to 10 decibels (dB) louder; and depending on their size and shape, ear trumpets could amplify by about 10 to 20 dB, with most of this in the range of 500 to 1000 Hertz (Hz). As the range of human speech is 300 to 3000 Hz, ear trumpets could only help people with mild hearing impairments.

Auricles and cornets were developed as an alternative to the ear trumpet. It was hoped that the device would be less observable on the user. Other hearing assistance devices used through the early part of the 1800s were bone conduction aids. Sounds are transmitted to the ear by vibrations in the air, but also by vibration of the bones in the skull. Bone conduction devices had been tested since the sixteenth century and were typically designed in two ways. The first design consisted of an apparatus held to the speaker's mouth or throat while the opposite end was held in the listener's teeth. The primary constraint of this model was the restriction on speaking distance and the number of persons involved in conversation. Allowing a greater distance from the speaker, the second design involved an instrument that collected sound energy by means of a flat transmitting surface from the surrounding air, something comparable to an acoustic fan. Inventors and physicians however were not satisfied with these devices, and by the end of the 1800s efforts were being made to produce a hearing aid more powerful and effective than the ear trumpet or bone conduction aids.

Electrical hearing aids were introduced just after 1900 and brought useful amplification to a wider audience. It is uncertain who invented the first electric hearing aid, but it may have been Miller Reese Hutchinson in 1898. This first electric hearing aid came to market in 1901 and used the transmitting potential of carbon in order to amplify the sound. There existed, however, substantial static noise and distortion caused by the carbon hearing aid. Such aids were also very large and impractical, but their effectiveness in amplifying sound surpassed any of the prior devices used. They offered the same amplification as ear trumpets, but covered a wider frequency range of 500 to 1800 Hz. All such devices consisted of the carbon microphone (technology borrowed from the telephone), processing unit, battery box and headpiece. Batteries often did not last more than a few hours and were very expensive. Later models with multiple microphones provided 25 to 30 dB of amplification, and the introduction of amplifiers in the 1920s increased the range to 45 to 50 dB.

The development of the vacuum tube hearing aids in the 1920s offered a device with an even greater amount of power and reduced slightly the size of the processing components. Wearable multipart hearing aids were then developed and used in the 1930s and 1940s. The only drawback with these aids was that batteries remained large and still far from invisible. A separate battery pack was needed to warm the vacuum tubes. In addition, most of the earlier large-size batteries did not last more than a day and had to be carried in special cases. This problem was resolved in 1947 with the invention of the transistor. As transistors got smaller, so did hearing aids, and concealment became an achievable and important goal.

By the late 1950s and early 1960s, models that were "at-the-ear" or "over-the-ear" combined a microphone with a battery and transistor in one unit and could amplify sounds within the range 400 to 4,000 Hz. These models were molded with custom-made ear tubes, and were easy to conceal

behind the ear or under the hair. In the 1970s, batteries became even smaller, allowing "in-the-canal" aids to fill the ear canal without anything worn outside the ear. By the late 1980s, advanced circuitry and lithium batteries made possible "in-the-ear-canal" units that could be concealed completely in the ear canal.

Since the 1990s, most manufacturers have produced four basic styles of hearing aids:

Behind-the-Ear (BTE)

The components are held in a case worn behind the ear and connected to a plastic earmold that fits inside the outer ear. Sound travels through the earmold into the ear. BTE aids are used by people of all ages for mild to profound hearing loss. Poorly fitting BTE earmolds may cause feedback, a whistle sound caused by the fit of the hearing aid or by the build up of earwax or fluid. However, BTE aids can be as sophisticated as smaller hearing aids. In fact, they can hold more circuitry and amplify sounds to a greater degree than in-the-ear types. BTE aids can be more durable than other types and a few are even waterproof.

In-the-Ear (ITE)

These devices house components in a custom-formed earmold that fits within the outer portion of the ear. ITE aids can accommodate added technical mechanisms such as a telecoil, a small magnetic coil contained in the hearing aid that improves sound transmission during telephone calls. ITE aids are used for mild to severe hearing loss but can be damaged by earwax and ear drainage. Their small size can cause adjustment problems and feedback. Usually, children do not wear them because the casings need to be replaced as the ear grows. However, its size and easy-to-use controls may be helpful for those with limited manual dexterity.

Canal Aids

These fit into the ear canal and are available in two sizes. In-the-canal (ITC) hearing aids are smaller still, with an earmold that fits down into the ear canal, and a smaller portion facing out into the outer ear. They are discreet, yet still visible within the outer ear. The ITC hearing aid can also be customized to fit the size and shape of the ear canal and is used for mild or moderately severe hearing loss. The newest generation of such hearing aids is those that fit completely in the canal (CIC). A CIC hearing aid is largely concealed in the ear canal and is used for mild to moderately severe hearing loss. In general, because of their small size, canal aids may be difficult for the user to adjust and remove, and they may not be able to hold additional devices such as a telecoil. Canal aids can also be damaged by earwax and ear drainage. They are not typically recommended for children.

Body Aids

Body aids are used by people with profound hearing loss. The aid is attached to a belt or a pocket and connected to the ear by a wire. Because of its large size, it can incorporate many signal processing options, but it is usually used only when other types of aids cannot be used.

Mechanisms of Action

The inside mechanisms of hearing aids vary among the different devices, even if they are of the same style. In general, every hearing aid is a miniature electronic circuitry encased in plastic. Every hearing aid consists of a microphone that picks up sound, an amplifier that boosts the sound, and a receiver that delivers the amplified sound into the ear. All the parts are powered by replaceable batteries. However, due to microprocessor or computer chip technology, late twentieth century hearing aids far surpass the simplicity of this description. All hearing aid styles described above use three basic types of circuitry:

1. An analog hearing aid works much the same way as traditional high-fidelity audio systems. Sound is picked up by the microphone and is then converted to electrical signals. Once sound is turned from acoustic to electrical signal, it is fed into the amplifier of the hearing aid. The sound is then amplified overall and sent to the receiver of the hearing aid, and finally to the user's ear. This type of sound processing has now been used over many years. The biggest drawback of analog processing is that the amplified sound is over the full frequency range of hearing, so low frequency (background noise) would "mask" high frequency (speech) sounds. To alleviate this problem, "potentiometers," which provide the ability to reduce or enhance the sounds needed by the end user, have been introduced to hearing aids to restore hearing to as "normal" a sound as possible. Analog circuitry is generally the least expensive.

2. Programmable hearing aids offer different (often customizable for the individual's audiogram) settings for different listening situations such as the office or home, or a noisy rock concert. Programmable devices enable the audiologist to fine tune the hearing aid using a computer. Potentiometers are built onto the circuit within the programmable hearing aid, rather than sitting externally on the hearing aid. The circuitry of analog and programmable hearing aids accommodates more than one setting. If the aid is equipped with a remote control device, the user can change the program to accommodate a given listening environment.
3. Hearing instruments incorporating digital signal processing (DSP) are widely known as digital hearing aids. The real difference between analog and digital hearing instruments is the ability of the DSP instruments to process more complex signal processing. Sound is still received into the DSP instrument by microphone and converted into "bits" of data. The circuitry within the "digital" hearing aid now acts like a very tiny computer. The computer can sample the data and far more accurately fine tune to each individual's requirements. These hearing aids are wearable minicomputers capable of monitoring the audiological environment. Digital hearing instruments compensate for different listening situations in a more flexible, accurate, and complex way than any analog circuit. Digital hearing aids use a microphone, receiver, battery, and computer chip. Digital circuitry can be used in all types of hearing aids and is typically the most expensive.

Implantable Hearing Aids

The latest twentieth century development in hearing aids is the implantable type. This type of hearing aid is placed under the skin surgically. Regarded as an extension of conventional hearing aid technology, an implantable hearing device is any electronic device completely or partially implanted to improve hearing. This allows for the hearing aid to be worn at or in the ear and still provide the user with a signal quality equal or superior to that of prior hearing aids.

Included in the category of implantable hearing aids is the cochlear implant, a prosthetic replacement for the inner ear or cochlea. In its most basic form, the cochlear implant is a transducer, which changes acoustic signals into electrical signals in order to stimulate the auditory nerve. The device is surgically implanted in the skull behind the ear, and it electronically stimulates the auditory nerve with small wires touching the cochlea. External parts of the device include a microphone, a speech processor (for converting sounds into electrical impulses), connecting cables, and a battery. Unlike a hearing aid, which just makes sounds louder, this device selects information in the speech signal and then produces a pattern of electrical pulses in the user's ear. It is impossible, however, to make sounds completely natural because a limited amount of electrodes are replacing the function of tens of thousands of hair cells in a normal hearing ear.

Hearing aid technology advanced greatly in the last few years of the twentieth century, thanks to the computer microchip and to digital circuitry. The majority of efforts in the research and further development of hearing aid technology is concentrated in the area of implantable hearing devices.

See also **Audiology, Implants and Surgery; Audiology, Testing**

PANOS DIAMANTOPOULOS

Further Reading

Berger, K.W. *The Hearing Aid: Its Operation and Development*, U.S. National Hearing Aid Society, 1984.
Dillon H. *Hearing AIDS*. Thieme, New York, 2001
Stephens, S.D.G. and Goodwin, J.C. Non-electric hearing aids to hearing: a short history. *Audiology,* 23, 215–240, 1984.
Valente, V. *Hearing AIDS: Standards, Options, and Limitations*. Thieme, New York, 2002

Useful Websites

Hearing Aid Manufacturers:
A&M Hearing: http://www.hearing-am.com
Amplifon: http://www.amplifon.it
Audio Controle: http://www.audiocontrole.com
Aurilink: http://www.aurilink.com
AVR: http://www.avrsono.com
Beltone: http://www.beltone.com
Bernafon: http://www.bernafon.ch
Cochlear: http://www.cochlear.com
Danavox: http://www.danovox.com
EarCraft Technologies: http://www.earcraft.com
Frye: http://www.frye.com
General Hearing Instruments: http://www.generalhearing.com
Magnatone: http://www.magnatone-hearing-aids.com
Micro-Tech: http://www.hearing-aid.com
Miracle-Ear: http://www.miracle-ear.com/main.asp
Oticon: http://www.oticon.com
Phonak: http://www.phonak.com
Phonic Ear: http://www.phonicear.com

Siemens: http://www.siemens-hearing.com
Starkey: http://www.starkey.com
Telex: http://www.telex.com/hearing/Hearinginstruments.htm
Unitron: http://www.unitronhearing.com
Viennatone: http://www.viennatone.com
Widex: http://www.widex.com

Audiology, Implants and Surgery

The cochlear implant is the first effective artificial sensory organ developed for humans. The modern cochlear implant has its origin in the late eighteenth century. Alessandro Volta, the Italian physicist whose work on electricity gave his name to a measure of electric current (volt), bravely conducted an experiment on his own ears to see what would happen. He attached two metal rods to activate an electric circuit and then inserted one rod into each ear. Volta reported that he felt something like a blow to his head, followed by the sound of boiling liquid. He did not repeat his experiment.

One of the first documented cases of an implant that successfully stimulated the auditory nerve took place in France in 1957. A French electrophysiologist named Andre Djourno teamed up with otolaryngologist Dr. Charles Eyries to implant a single electrode directly on the auditory nerve of a deaf man and connect it to a crude signal generator. The man described sounds like crickets and was able to use his improved awareness of speech rhythm to lip-read more easily.

In the 1960s, Dr. William House in Los Angeles started the movement toward commercial development with a single-channel, single-electrode device that was later known as the 3M/House device. In 1964, Dr. Blair Simmons at Stanford University placed the first multichannel implant in a human. When he submitted a paper on the subject for a presentation at the American Otological Society in the following year, however, it was rejected as too controversial. By the 1980s and 1990s, cochlear implants had become generally accepted as a safe and effective method of treating sensorineural deafness in adults and children.

A 1995 U.S. National Institutes of Health Consensus statement on cochlear implants in adults and children concluded that "a majority of those [adult] individuals with the latest speech processors for their implants will score above 80 percent correct on high-context sentences even without visual cues." The development of this technology represents one of the most rapid advances ever seen in medical technology; by the end of the twentieth century, there were about 45,000 users of cochlear implants around the world, almost half of them children. Deaf children who receive implants early enough can now be expected to acquire spoken language and the ability to function well in a hearing milieu.

While hearing aids can be implanted or non-implanted and simply amplify incoming sound, a cochlear implant is always surgically implanted and directly stimulates the auditory nerve. Candidates for cochlear implants are typically those with profound or severe sensorineural deafness in both ears. They normally perform signifi-

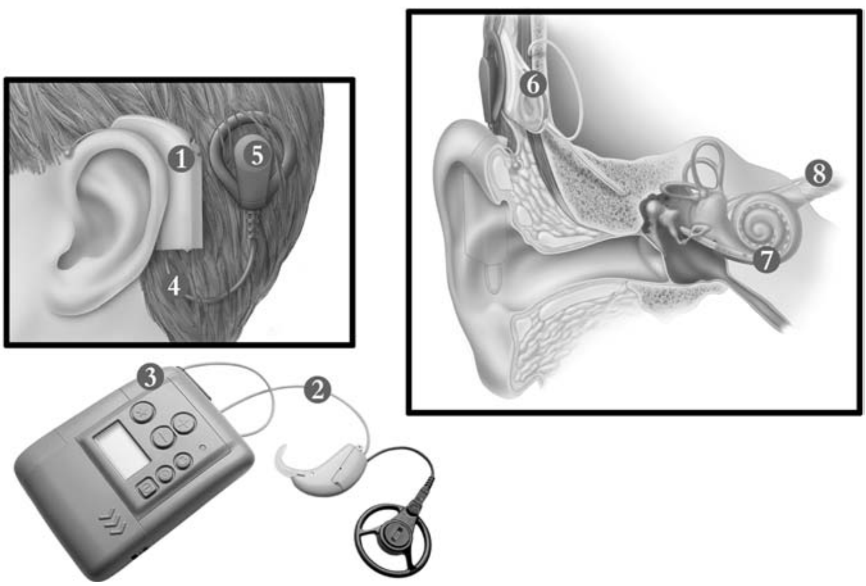


Figure 3. Nucleus® 24 Contour™ cochlear implant system. 1. Sounds are picked up by the small, directional microphone located in the headset at the ear. 2. A thin cord carries the sound from the microphone to the speech processor, a powerful miniaturized computer. 3. The speech processor filters, analyzes and digitizes the sound into coded signals. 4. The coded signals are sent from the speech processor to the transmitting coil. 5. The transmitting coil sends the coded signals as FM radio signals to the cochlear implant under the skin. 6. The cochlear implant delivers the appropriate electrical energy to the array of electrodes which has been inserted into the cochlea. 7. The electrodes along the array stimulate the remaining auditory nerve fibers in the cochlea. 8. The resulting electrical sound information is sent through the auditory system to the brain for interpretation.

cantly better with a cochlear implant than they would with a hearing aid.

The internal parts of a cochlear implant consist of two parts—a surgically inserted electrode array threaded through a hole drilled in the inner ear and inserted into the cochlea, and a receiver/stimulator magnet that is placed over the mastoid bone beneath the skin and behind the ear. The external parts consist of a digital processor, a microphone, and a transmitter. The transmitter sits on the head behind the ear and is held securely in place by its magnet, which is attracted to the magnet of the receiver/stimulator under the skin. The appearance, size, and attributes of components vary depending on the manufacturer and the date of manufacture. However, the general way in which the systems work is the same in all devices.

The cochlear implant works when the microphone sitting on the wearer's head picks up sounds in the environment and sends them to the digital processor by way of a thin cable. The processor uses an internal program that has been customized to the thresholds and comfort levels of the patient to rapidly convert the sounds into electrical codes. These codes indicate which electrodes should be stimulated and how (for example, at what intensity). The coded signals pass back along the cable to the flat transmitter resting on the wearer's head. The transmitter then sends the information across the skin as radio frequency signals to the receiver/stimulator under the skin. There, the signals are decoded and sent on to the electrode array within the inner ear to stimulate the appropriate electrodes and thence the adjacent auditory nerve endings. The signals are passed along the auditory nerve to the brain, where they are interpreted finally as sound. The entire cycle from receipt of sound by the microphone to deciphering it in the brain is rapid: in the slowest systems, the entire cycle takes place 250 times per second; in the fastest systems, the cycle can take place up to 1,000,000 times per second.

Candidates for cochlear implants require an intact auditory nerve. For those who do not have an intact auditory nerve, an auditory brain stem implant (ABI) is possible. This was first developed in the 1970s at the House Ear Institute in the U.S. The ABI is based on the technology of the cochlear implant and works in the same way, except the implanted portion is placed directly over the cochlear nucleus of the brain stem instead of within the inner ear, thus bypassing the ear completely. By 2000, approximately 300 individuals had received an ABI. People generally do not hear as well with ABIs as with cochlear implants.

There is a learning period involved in making sense of the new sounds delivered by either type of implant. Some people learn more quickly than others, and some achieve greater success in understanding speech without lip reading than others. Not all the factors affecting success are well understood, although the length of time the person has been deaf is considered the main factor.

There are three major companies developing the technology of cochlear and auditory brainstem implants: Cochlear Limited (based in Australia), Advanced Bionics (based in the U.S.), and MED-EL (based in Austria). The oldest and largest is the Australian company.

The adaptation of individuals to this technology, which was deemed quackery in the 1960s, is now teaching researchers much about how normal hearing works in those who are not deaf. Two major challenges faced developers of the technology at the beginning of the twenty-first century. First, there is a need to improve the ability of patients to hear with the devices in noisy environments. Second, the programming strategies used in the processors, as well as the fine-tuning techniques to customize them, need to become more sophisticated to maximize the individual's potential success and reduce the current variability in performance.

See also **Audiology, Hearing Aids; Audiology, Testing.**

BEVERLY BIDERMAN

Further Reading

Biderman, B. *Wired for Sound: A Journey into Hearing*. Trifolium Books, Toronto, 1998. An inside account of learning to hear with a cochlear implant. Appendices include extensive endnote references to the literature on the subject, as well as listings of resources on deafness and cochlear implants.

Clark, G.M., Cowan, R.S.C. and Dowell, R.C., Eds. *Cochlear Implantation for Infants and Children: Advances*. Singular Publishing, San Diego, 1997. Articles by the Melbourne Australia team of researchers and surgeons. The chapter, "Ethical Issues" is a thorough treatment of the ethics of cochlear implantation in children, a controversial issue.

Clark, G.M., Tong, Y.C. and Patrick, J.F., Eds. *Cochlear Prostheses*. Churchill Livingstone, Edinburgh, 1990. Excellent material on the development of the Melbourne prototype which was later commercially developed as the Nucleus device by Cochlear Limited in Australia.

Niparko, J., Kirk, K., Mellon, N., Robbins, A., Tucci, D. and Wilson, B. *Cochlear Implants: Principles and Practices*. Lippincott, Williams & Wilkins, Philadelphia, 1999. The clinical practices related to

cochlear implantation from selection of candidates to techniques of device placement and activation.
Salvi, R.J., Henderson, D., Fiorino, F. and Colletti, V., Eds. *Auditory System Plasticity and Regeneration*. Thieme, New York, 1996.A collection of papers discussing the status of research into auditory plasticity, including papers on ear hair cell regeneration, the effects of deafness and electrical stimulation on the auditory system, and a discussion of cochlear implant performance over time as an indicator of auditory system plasticity.
Schindler, R.A. and Merzenich, M.A., Eds. *Cochlear Implants*. Raven, New York, 1985. Of great historical importance, this publication is based on the proceedings of the Tenth Anniversary Conference on Cochlear Implants, June 22–24, 1983 at the University of California, San Francisco. The papers by Drs. House, Simmons, and Michelson, early pioneers of the technology, are especially interesting.

Audiology, Testing

The *Archives of Paediatrics* of 1896 reported that 30 percent of the large number of children with poor educational and mental development in Germany and America had deficient hearing power. At the close of the nineteenth century, methods for measuring hearing included observing the patient's ability to detect sounds of a ticking watch, Galton's whistle, the spoken voice, and tuning forks. Half a century earlier in Germany, Hermann Ludwig Ferdinand von Helmholtz investigated hearing physiology and Adolf Rinne used tuning forks to diagnose types of deafness. Galton's whistle emitted accurate variations in tones without loudness control and allowed exploration of hearing ability; Adam Politzer's acoumeter produced loud controlled clicks for detecting severe deafness. These principles of intensity and frequency measurement have become integral features of modern audiometers.

Induction coil technology used in Alexander Graham Bell's telephone (1877) formed the basis of many sound generators. In 1914, A. Stefanini (Italy) produced an alternating current generator with a complete range of tones, and five years later Dr. Lee W. Dean and C. C. Bunch of the U.S. produced a clinical "pitch range audiometer." Prerecorded speech testing using the Edison phonograph was developed in the first decade of the twentieth century and was employed in mass testing of school children in 1927.

World War I brought about major advances in electro-acoustics in the U.S., Britain, and Germany with a revolution in audiometer construction. In 1922 the Western Electric Company (U.S.) manufactured the 1A Fowler machine using the new thermionic vacuum tubes to deliver a full range of pure tones and intensities. Later models restricted the frequencies to the speech range, reduced their size, and added bone conduction.

In 1933 the Hearing Tests Committee in the U.K. debated the problems of quantifying hearing loss with tuning forks because their frequency and amplitude depended on the mode of use. Despite the availability of electric audiometers at the time, standardization of tuning fork manufacture and strict method of use was recommended, along with the introduction of graphic documentation audiography.

Concern abroad about a standard of intensity of sound, or loudness, led to the American Standards Association adopting the bel, named after Alexander G. Bell and already used as a measurement in the acoustic industry. The decibel (dB) was adopted as a unit of measurement of the relative intensity of sounds in the U.S., Britain, and then Germany. International standards were established in 1964.

Subjective Hearing Testing

Several tests for hearing involve the subject's response to types of sounds in a controlled setting. Tests are designed to detect conductive loss, which results from obstruction and disease of the outer or middle ear, and sensorineural (perceptive) loss, caused by lesions of the inner ear (cochlea) or auditory nerve (nerve of hearing) and its connections to the brain stem. Tuning fork tests compare responses to sound transmitted through the normal route of air conduction (AC) as well as bone conduction (BC) through the skull bone to the cochlear shell. Conductive loss reduces air conduction, and perceptive disorders reduce bone conduction. The Rinne test (Figure 4) is positive when AC > BC and negative when AC < BC.

Pure tone audiometry (PTA) is a subjective measurement of AC/BC thresholds using headphones. Responses are displayed graphically. Frequencies cover 250–8,000 Hz, the range for speech understanding, and are plotted against intensities of –10 to 120 decibels, hearing level (dBHL). The internationally agreed normal threshold for hearing is 0 dB and is calibrated from healthy young adults. Classification of the degree of hearing loss is based on pure tone audiometry of each ear.

Speech audiometry tests the subject's response to phonetically balanced words at different intensities. The test is used to evaluate hearing discrimination for fitting hearing aids and for cochlear implantation.

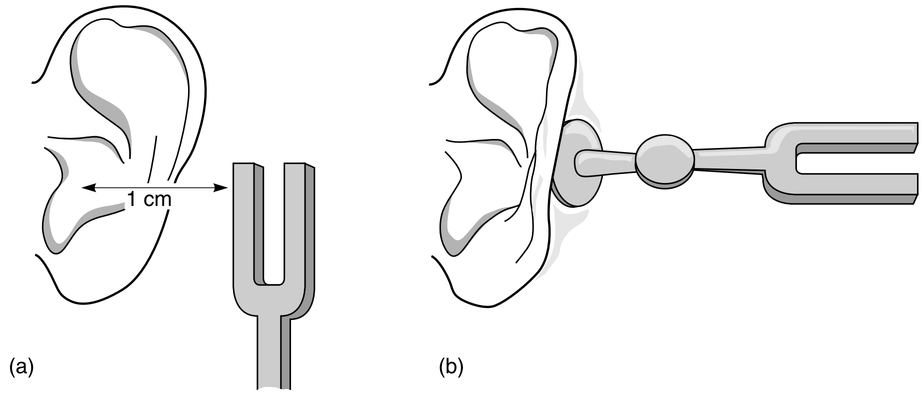


Figure 4. Rinne test using a tuning fork: (a) air conduction; (b) bone conduction. [*Source: Ludman H. and Kemp, D. T. Basic Acoustics and Hearing Tests, in Diseases of the Ear, 6th edn., Ludman, H. and Wright, T., Eds. Edward Arnold, London, 1998. Reprinted by permission of Hodder Arnold.*]

Behavioral Testing

Around 1900, neurophysiology gained importance with the work of Charles Scott Sherrington in the U.K. and Ivan Petrovich Pavlov in Russia who were exploring innate and conditioned reflexes. Over the next few decades, interest spread beyond Europe to Japan and the U.S. Clinicians began to use these subjective behavioral responses along with various noise makers to identify hearing loss in infants and children. The year 1944 was an important landmark for the introduction of a developmental distraction test researched by Irene R. Ewing and Alexander W. G. Ewing at Manchester University. This first standard preschool screen became a national procedure in Britain in 1950 and was later adopted in Europe, Israel, and the U.S. Observations of head turns to meaningful low- and high-frequency sounds in infants about 8 months old were made with one person distracting and another testing. This screening method, although still in use, failed to detect moderate deafness.

Automated behavioral screening to detect body movements in response to sounds represented a transition period in the mid-1970s. Two tests, the American Crib-o-gram (1973) and Bennett's Auditory Cradle (London 1979), both proved inefficient.

Objective Testing

In 1924 German psychiatrist Dr. Hans Berger recorded brain waves with surface cranial electrodes. Electroencephalography (EEG) was used fifteen years later by P. A. Davis in the U.S. to record electrical potentials from the brain with sound stimulation. In the journal *Science* in 1930, Ernest Weaver and Charles Bray, also in the U.S., confirmed the production of electrical potentials in the inner ear to sound stimulation in cats. In New York in 1960, action potentials were recorded with electrode placement through a human middle ear directly onto the inner ear. Electrocochleography (ECoG) was used for diagnosis but required anesthesia, especially for children, and was replaced by noninvasive auditory brainstem response audiometry (ABR).

Signal-averaging computerization in the 1950s enhanced the recordings of synchronous neural discharges generated within the auditory nerve pathway with EEG on repetitive acoustic stimulation. Research in the later 1960s in Israel and the U.S. identified amplified electrical potentials more easily in cats and humans with improved computerization to extract background noises. Peaked waves I to VII represented various component activities along the auditory route, wave V being the most diagnostic (Figure 5). ABR was thoroughly investigated in the U.S. and was validated in 1974 for testing high-risk infants. An automated model, automated auditory brain stem response (AABR) was manufactured in the U.S. during the next ten years and proved to be an affordable, noninvasive screening tool for babies. The device used short broadband clicks to elicit responses in the 2000 to 4000 Hz speech recognition range and compared the results with a template algorithm of age-related normal responses.

In 1978 David Kemp in London researched earlier reports of sound reemissions from the ear and redefined outer hair-cell function (inner ear workings). These sensors "amplified" and converted the mechanical sound waves traveling along the cochlear fluid duct, which sharpened and strengthened each individual sound to stimulate a specific anatomical site along the duct. The excess energy produced by this process was reflected as "echoes" and could be collected with Kemp's miniaturized ear canal microphone. These otoacoustic emissions (OAE) were ana-

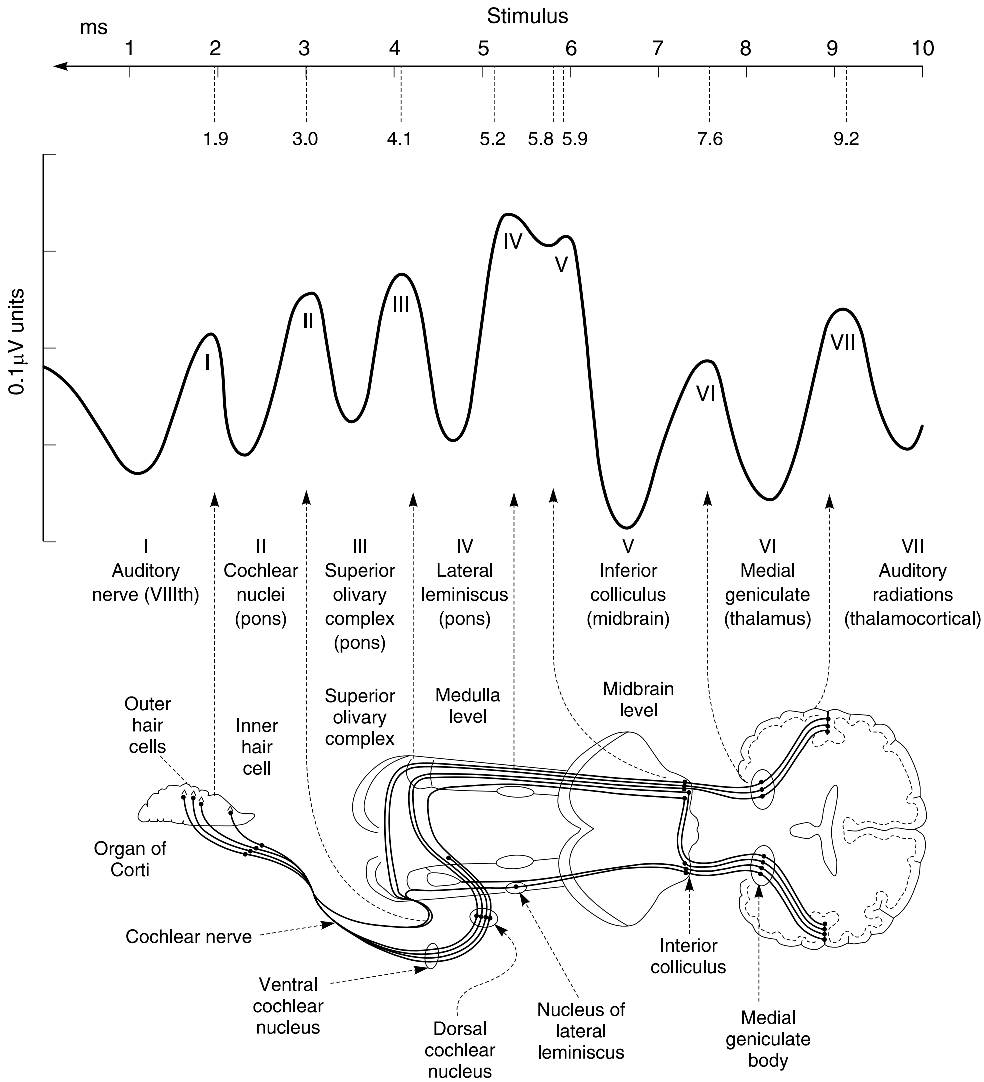


Figure 5. Normal brain stem responses and anatomical correlates of waves I to VII in auditory brain stem response (ABR).
[*Source: Ludman H. and Kemp, D. T. Basic Acoustics and Hearing Tests, in Diseases of the Ear, 6th edn., Ludman, H. and Wright, T., Eds. Edward Arnold, London, 1998. Reprinted by permission of Hodder Arnold.*]

lyzed by computerization according to specific frequency responses, thus providing an extremely sensitive apparatus for detecting cochlea hair-cell damage (Figure 6). After trials in France it was widely accepted as an efficient and safe testing tool.

Infant Screening

The OAE and AABR physiologic tests have proved successful in detecting permanent deafness in infants and consequently in encouraging early intervention with hearing aids and communication strategies. Universal newborn hearing screening programs have been established internationally.

See also **Audiology, Implants and Surgery; Audiology, Hearing Aids; Electronics; Valves or Vacuum Tubes**

DAVID ROSE

Further Reading

Bender, R.E. *The Conquest of Deafness*. Interstate Printers & Publishers, Denville, IL, 1981.

Campbell, K. *Essential Audiology for Physicians*. Singular Publishing Group, San Diego, 1988.

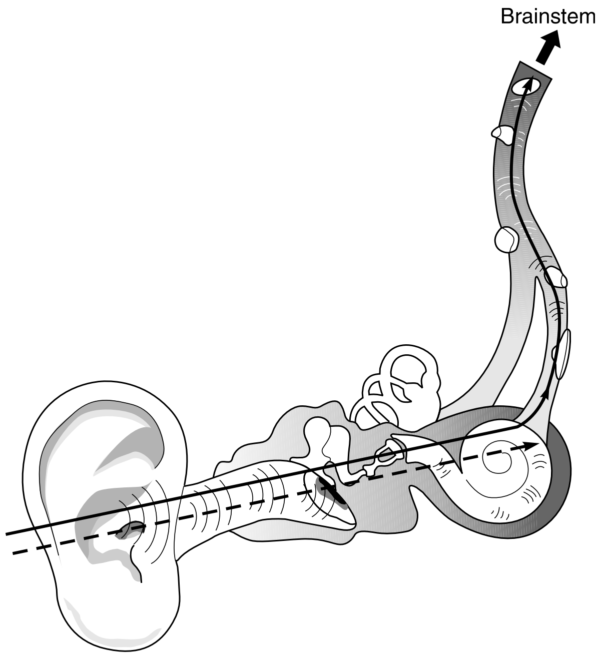


Figure 6. Screeners use automated auditory brain stem response (AABR) technology to test the entire hearing pathway from the ear to the brain stem. Otoacoustic emissions (OAE) test hearing from the outer ear to the cochlea only.
[*Used with permission of Natus U.K.*]

Ewing, I.R., and Ewing, A.W.G. *The Handicap of Deafness*. Longmans, Green & Co., London, 1938.

Ewing, A.W.G. The education of the deaf: history of the Department of Education of the Deaf, University of Manchester, 1919–1955. *Br. J. Educ. Stud.* 4, 2, 1956.

McGrew R.E. *Encyclopaedia of Medical History*. Macmillan, London, 1985.

Sebastion A. *A Dictionary of the History of Medicine*. Parthenon, Carnforth, 1999.

Automobiles

An automobile is commonly defined as a self-propelled, minimum three-wheeled vehicle for passenger or freight transport on ordinary roads. Today's cars are commonly driven by an internal combustion engine using volatile fuel.

Without any doubt, the motor vehicle is among the technological developments of the twentieth century with the most far-reaching social and economic consequences. Automobile production became a major industry in many industrialized countries and had a decisive influence on the development of modern mass production technologies. Furthermore, by deeply transforming the living habits of people, automobiles became an integral part of modern western culture.

At the turn of the twentieth century, these developments were not at all foreseeable. Many prototype solutions for the various subsystems of automobiles had already been developed, but all in all there was remarkable uncertainty about how a proper automobile should be designed. Commercial car production was well established, but the automobiles were still individually crafted and very expensive.

The twentieth century began with a real technical break in 1901 when the first Mercedes racing-car of the German Daimler–Motoren–Gesellschaft put an end to the ''horseless carriage'' period of the late nineteenth century. The new vehicle incorporated a host of improved features in a general concept that was exemplary: honeycomb radiator mounted right at the front of the car with directly attached engine hood (bonnet); longitudinal four-cylinder in-line engine with the gearbox placed directly behind it; drive to the rear axle; lightweight pressed-steel frame with large wheelbase. The international diffusion of this future standard layout characterized the technical development in the first decade of the twentieth century.

Even the most famous car of these years, the 1908 Model T of the Ford Motor Company, followed this concept. With the Model T, Henry Ford succeeded in building a reliable, high quality, and relatively low-cost car with an initial price of $950. However, even in the U.S., with its comparatively high incomes, a further price reduction was needed to create a real mass market for automobiles. Ford and his staff accomplished this through a fundamental modification of car production technology and organization. Continuous flow production, widespread use of single-purpose machine tools and, from 1913 onward, a step-by-step changeover to assembly line production revolutionized car manufacturing. With Ford's system of rigid mass production, automobiles became affordable consumer goods for middle class Americans. In 1916 the price of the Model T touring car was $360; in 1927 it was $290.

In the years before World War I, cars in general became more reliable, comfortable, and suitable for everyday purposes. In Europe, unlike the U.S., automobile purchases remained limited to a small group of wealthy customers. Only the first steps were taken toward the professional use of cars, for example by physicians.

After 1918 in the U.S., increasing mass motoring had a strong influence on automobile design. The general change from open ''touring'' cars to larger and more powerful closed limousines for day-to-day use typified the development. In the 1930s,

U.S. cars became remarkably more comfortable with new features such as elastic engine support, automatic transmission, and overdrive, but the conventional drive layout (front engine, rear wheel drive, rigid axles with leaf springs) remained unchanged.

The European automobile industry recovered only slowly after World War I. The still comparatively low incomes and high operating costs prevented the emergence of a real mass market for automobiles. European companies therefore took over the American mass production system only piecemeal and with hesitation. On the other hand the higher flexibility of the relative small European manufacturers allowed them to be much more innovative. In the 1930s, the increasing road speed of European cars forced the industry to overcome the restraints of conventional car design: swing and floating axles, front wheel drive, and aerodynamic bodies entered series production.

During World War II, the development of new models came to a halt in all car-producing countries. After the war, production resumed on the basis of prewar models. In the 1950s, manufacturers and customers in the U.S. favored comfortable, powerful, ‘‘gas guzzling’’ cars with prestigious chromium-plated bodies still designed on the conventional drive layout. Outside the U.S. these expensive models could hardly be sold.

In Europe, the 1950s were years of rapid change. All enveloping bodies and automatic transmissions were adopted from the U.S., and self-supporting bodies, fuel injection, radial-ply tires, hydropneumatic suspension systems, and servo components, improved the performance, handling, and safety of automobiles. According to the market demand, European producers focused on small and medium sized cars. Based on models such as the Morris Minor, the Citroen 2 CV, the Fiat 600, and the Volkswagen Beetle, Western Europe now finally experienced its own mass production boom in automobiles (see Table 2.)

In the 1960s the dark side of automobile traffic became more and more obvious and began, first in the U.S., to generate public debate on safety and environmental risks. The first energy crisis of 1973–1974 intensified the debate, shifting the focus to fuel consumption. Air pollution, safety, and fuel economy as public policy issues and the resulting governmental regulations exerted great pressure on the international automobile industry. A new standard type of smaller cars with front wheel drive, front mounted transverse engine, McPherson front suspension, and fastback design was developed and became dominant on the market. The BLMC-Mini of 1959 was already a decisive ancestor of this new style, but it was the 1974 Volkswagen Rabbit (Golf) that was its first proper representative. Fuel-saving Otto and especially diesel engines gained an increasing market share. Safety components such as disk brakes, seat belts, and airbags were introduced and became standard features. With the introduction of catalytic converters to treat exhaust emissions, the ecological damage through growing car traffic was at least limited.

Table 2 Automobiles per 1,000 Inhabitants in Selected Countries 1950–1995.

	1950	1960	1970	1979	1986	1995
U.S.	260	340	430	527	681	723
France	40	110	240	355	435	521
U.K.	50	110	210	256	393	472
Germany[1]	40	90	230	357	456	540
Italy	15	30	190	303	411	569
Japan	–	5	80	185	381	520

[1] 1950–1990 West Germany, 1995 united Germany.

In the 1970s Japan became a new global competitor on the international scene. With their comparatively small, inexpensive, and fuel-saving models, Japanese auto manufacturers met the changed consumer priorities and got a firm footing on the North American and Western European market. The international adoption of Japanese production methods (lean production, just-in-time system, robot-based flexible mass production) caused a second revolution of car manufacturing.

In the 1980s and 90s environmental, fuel economy, and safety concerns still exerted a major influence, and electronic components became a key technology in automobile engineering. Microelectronic controls have not only optimized the motor management (ignition and fuel injection) but also the transmission and suspension, steering, and braking systems. The trend now is toward central electronic control units that harmonize the various subsystems of which a car consists.

Influenced by the pollution debate and with a view to the fact that the global mineral oil deposits will inevitably be drained one day, different alternatives to the conventional car have been developed in the last 30 years. At present, vehicles

with hydrogen-powered Otto engines and particularly electric-powered cars with fuel cells are regarded as the most promising solutions. However, at least in the medium term, the automobile with fuel-powered Otto or diesel engine will remain prevalent. The fuel economy of the individual car will be further reduced by consequent lightweight design and an even broader application of electronic controls. Due to the ongoing process of increasing automobile use globally however, overall emissions will nevertheless continue rising.

See also **Automobiles, Electric; Automobiles, Hybrid; Automobiles, Internal Combustion; Internal Combustion Piston Engine**

REINHOLD BAUER

Further Reading

Bardue, J.P. *et al. The Automobile Revolution: The Impact of an Industry*. University of North Carolina Press, Chapel Hill, 1982.

Flink, J.J., *The Automobile Age*. MIT Press, Cambridge, MA, 1988.

Mondt, J.R. *Cleaner Cars, The History and Technology of Emission Control since the 1960s*. Society of Automotive Engineers, Warrendale, 2000.

Newcomb, T.P. and Spurr, R.T. *A Technical History of the Motor Car*. Adam Hilger, Bristol, 1989.

Rae, J.B., *The American Automobile Industry*. MIT Press, Cambridge, MA, 1990.

Rubinstein, J.M. *Making and Selling Cars: Innovation and Change in the U.S. Automotive Industry*. Johns Hopkins University Press, Baltimore, 2001.

Volti, R.A. Century of automobility. *Technol. Cult*., 37, 663–685, 1996.

Automobiles, Electric

Since its development in the 1890s, the electric automobile has always been the car of tomorrow. Many of the earliest motor vehicles employed electric technology, and thousands of electric vehicles provided satisfactory transport service in the U.S. and Europe over the course of the twentieth century. Never as incapable as its critics claimed, neither was the electric automobile ever destined to become the "universal" vehicle of choice for the typical driver. In specific places and times and for select applications, the electric vehicle excelled, but it has never managed to live up to its lofty expectations.

Prior to 1903, the very soul of the automobile was at stake. Three factors led to the initial failure of the electric vehicle. First, early consumers played an outsized role in defining the overall shape of the technology. Wealthy elite males in search of the "sporting life" sought a technology that offered excitement and enabled exurban touring. Although many wealthy families also "stabled" an electric automobile along with their other cars and horses and used it for local transport, the electric vehicle was too practical and domesticated to satisfy these deeper symbolic needs. Successful applications of electric vehicles focused on commercial needs. Electric taxicabs operated continuously in New York City from 1897 to 1912, but these operated on a traditional, horse-based livery model of transport service. The vehicles were leased and not owner operated.

Second, exurban touring required infrastructure that favored internal combustion automobiles. In addition to roads and service facilities, which were decidedly less conducive to electric touring, access to electricity was inherently problematic. Regardless of the absolute range and the technical challenge of charging batteries in remote locations, symbolically the electric car was always "tethered to a wire." The presence of electric service was evidence of civilization and therefore the very thing that early automobilists sought to escape.

Finally, electric technology suffered from unmet expectations, while internal combustion managed to greatly exceed its initial promise. Following the electrification of lighting and street railways, many observers expected electricity to soon make it "off the rails," whereas few saw the internal combustion engine as capable of powering a transportation revolution. By 1906, the tables had turned. Rapid advances in materials and machining of gasoline engines enabled higher compression and resulted in dramatic increases in both power-to-weight ratio and reliability. The universal electric automobile had failed, and mass motorization via internal combustion had begun.

Until the end of World War I, electric vehicles continued to prosper, but always in niche markets or as parts of larger vehicle fleets. Where fire danger limited the operation of combustion engines, electric stevedores replaced horses on docks and train platforms. Electric materials handling vehicles were used inside factories and storehouses where delicate merchandise could not be exposed to dirty exhaust fumes. Local delivery service was the most durable and persistent niche market. Through the 1920s in the U.S. and well into the post-World War II era in Britain, fleets of electric vehicles delivered everything from mail to milk. The unique attributes of the electric car—quiet, reliable, economical, and capable of frequent starts and stops—could not be matched.

During the middle decades of the century, electric vehicles were limited to materials handling

and other niche applications and all but disappeared from public view. In the mid-1960s, the electric automobile again emerged as the car of the future. Growing public awareness of the increasing costs of unbridled expansion of the gasoline automobile system and industry inertia prompted policy makers to look anew at the general purpose electric car. In the United States, after 1973, energy independence provided an additional argument in support of the electric vehicle, ultimately leading to the passage of the Electric and Hybrid Vehicle Research Development and Demonstration Act (1976). However, internal combustion had a 75-year head start, and expectations of range and performance had coevolved with suburbanization. Despite an infusion of government research funds, the electric vehicle proved incapable of displacing the established internal combustion standard.

This cycle repeated itself again in the 1990s. In January 1990, General Motors unveiled the Impact, a prototype electric automobile that offered greatly increased performance and range through the use of advanced materials and design. Later that year, the California Air Resources Board passed regulations requiring that a fraction of all vehicles sold in the state must be Zero Emissions Vehicles (ZEVs). Only electric vehicles met the initial criteria for ZEVs, and a decade-long battle ensued between the industry and state and federal regulators. As a byproduct, small numbers of electric cars and trucks were sold or leased by mainstream manufacturers like Chrysler, Ford, General Motors, Honda, Nissan, and Toyota. Enthusiasts adored the vehicles, but overall sales were disappointing, and industry support for the vehicles was tepid at best. Similar stories unfolded in various Western European settings. Tiny minorities were well served by the electric option, but not surprisingly the technology was still incapable of competing head-to-head with internal combustion. Gradually, researchers and policy makers again abandoned the electric vehicle, instead pinning hopes on other technological options, including hybrid electric vehicles (reintroduced by Toyota in Japan in late 1996 and in the U.S. in 2000) and the oft-promised hydrogen fuel cell.

Throughout the history of the electric automobile, one popular misconception stands out: Michael Schiffer has called it the ''better battery bugaboo,'' the idea that the electric vehicle failed entirely on account of the limited range of its batteries. While storage batteries have their limitations, range only became a binding constraint after internal combustion emerged as the leading propulsion technology. Before 1903, electric vehicle taxicab operators like the Electric Vehicle Company in New York City struggled with the cost of operation rather than range *per se*. Batteries were short-lived and therefore expensive, as were tires, but exchangeable batteries enabled vehicles to be refueled quickly, often while patrons waited in a lounge at the cab station. As noted herein, factors beyond the battery were at least as important, if not more so, in determining the ultimate fate of the electric car in the twentieth century.

See also **Automobiles; Automobiles, Hybrid; Automobiles, Internal Combustion; Batteries, Primary and Secondary; Fuel Cells**

DAVID A. KIRSCH

Further Reading

Kemp, R. *et al. Experimenting for Sustainable Transport: The Approach of Strategic Niche Management*. Spon Press, London, 2002.
Kirsch, D.A., *The Electric Vehicle and the Burden of History*. Rutgers University Press, New Brunswick, NJ, 2000.
Kirsch, D.A. and Mom, G. Visions of transportation: the EVC and the transition from service to product based mobility. *Bus. Hist. Rev.*, 76, 1, 75–110, 2002.
Mom, G. *The Electric Vehicle: Technology and Expectations in the Automobile Age*. Johns Hopkins University Press, Baltimore, 2004.
Mom, G. and Kirsch, D.A. Technologies in tension: horses, electric trucks and the motorization of American cities, 1900–1925. *Technol. and Cult.*, 42, 3, 489–518, 2001.
Schiffer, M. *Taking Charger: The Electric Automobile in America*. Smithsonian Institution Press, Washington D.C., 1994.
Volti, R. Why internal combustion? *Am. Herit. of Invent. Technol.*, 6, 2, 42–47, 1990.

Automobiles, Hybrid

In technology, as in biology, a hybrid is the result of ''cross-fertilization,'' in this case referring to the application of technologies to produce a similar yet slightly different entity. Recent research in the history of automotive technology shows that hybridization has been much more common than previously thought. Thus, the automobile itself can be viewed as a hybrid with a century-long history of crossover phenomena from electrical engineering to mechanical engineering that resulted in an ''electrified gasoline car.''

The term hybrid, however, is generally reserved for combinations of *propulsion systems* in automobiles. Most common in the history of the automobile is the thermoelectric hybrid, mainly a combination of the internal combustion engine (gasoline or diesel) and an electric propulsion

system (electric motor, battery set). Thermomechanical hybrids are possible when a combustion engine is combined with a flywheel system in which part of the kinetic energy during braking can be stored and released the moment this energy is needed, for instance for acceleration from standstill. Similarly, thermohydraulic hybrids combine combustion engines with a hydraulic energy storage system (a pump and a hydraulic accumulator). Electroelectric hybrids are also known; in these cases the actual propulsion is done by one electric motor, but the energy supply is a combination of battery storage and a supply from overhead trolley wires. Combinations of trolley systems and mechanical flywheel storage systems have also been built. Viewed from this perspective, the automobile as we know it at the end of the twentieth century is but one case among many possibilities.

Thermoelectric hybrids are nearly as old as automotive technology. Before 1900, the Belgian automobile producer Henri Pieper developed a car that was equipped with an electromagetically controlled carburetor. His patents were later bought by car manufacturers like Siemens–Schuckert (Germany), Daimler (Coventry, U.K.) and the French Société Générale d'Automobiles électro-Mécaniques (GEM). In 1908 the latter company proposed a Pieper-like hybrid called "Automobile Synthesis." At about the same time, German battery producer AFA (now Varta AG) bought a Pieper to develop a special battery for hybrid car applications. Another famous hybrid vehicle builder was the French electrical engineer Louis Kriéger. He started hybrid development in 1902 and produced a car he drove during the rallye from Paris to Florence a year later. In 1904 his hybrid was the sensation of the Paris automobile show. In 1906 he conceived a drive train based on an electric propulsion system and a gas turbine, and in the same year he developed a hybrid taxicab, 100 of which were intended to be built.

In Austria, Lohner built 52 hybrids between 1898 and 1910, designed by electrical engineer Ferdinand Porsche. These cars were later sold by Daimler, Germany, which founded a separate company for this purpose, Société Mercedes Mixte. In Germany several local fire companies built thermoelectric fire engines, some of these a combination of an electric motor with batteries and a steam engine. In this configuration, the electric drive system was meant for quick starting and for use during the first few kilometers of the trip. After ten minutes, when kettle pressure had built up, the steam engine took over to propel the truck to the fire location. All in all however, no more than a hundred or so hybrids were sold in Europe before World War I. In the U.S., there was even less hybrid construction activity during this period, the most famous being the Woods Dual Power, which was produced during the war.

Hybrids were supposed to combine the advantages of two systems while avoiding their disadvantages. For instance, because the thermal element in the hybrid system was often used (in combination with the electric motor, which for this purpose had to be repolarized to become an electricity generator) to supply a part of the stored electricity, the battery set in a hybrid tended to be smaller. It was lighter than that in a full-blown electric motor where all the energy for a trip had to be stored in the batteries before the start of the trip. In most cases the combination of systems led to a more complex and expensive construction, jeopardizing state-of-the-art reliability standards, and complicated control problems, which would only be overcome with the emergence of postwar automotive electronics. Also, despite the lighter battery, the total drive train became heavier. For this reason hybrid alternatives were especially popular among producers of heavy vehicles such as buses and trucks in which the relative importance of the drive train weight is less. Well-known examples in this respect are the brands Fisher (U.S.), Thornicroft (U.K.) and Faun (Germany).

The popularity of hybrid propulsion systems among engineers was not only, and according to some analysts not primarily, the result of technical considerations. During the first quarter century of automotive history, when the struggle between proponents of steam, electric, and gasoline engine propulsion was not yet over, hybridization often functioned as a strategic and social compromise as well. This was very clear in the case of the German fire engine community before World War I. A fierce controversy raged over the apparent unreliability of the gas combustion engine, but the proponents of electric drive trains, who boasted that electric drive trains guaranteed quick starting, high acceleration, high reliability, and no danger of fuel explosions in the neighborhood of fires, were not strong enough to monopolize the field. Several fire officials then opted for a hybrid drive, combining the advantages of electric with the advantages of the combustion engine (primarily a greater driving range), but they encountered heavy resistance from a combination of both other fire officials and the established automobile industry. Nevertheless, in 1910 the German fire engine fleet included about 15 heavy hybrids.

As with the electric alternative, hybrid automobiles experienced a revival during the last quarter of the twentieth century. This resulted in at least one commercially available hybrid automobile, the Toyota Prius. During this period, the issue of energy consumption played a role as well. Heavily subsidized by local, regional, and federal governments in Europe, Japan, and the U.S., hybrid projects used new light materials such as magnesium, plastics, and carbon fibers; and sophisticated electronic control systems (borrowed from related industries such as aerospace and information and communication technology) to enable very energy efficient solutions, initially to the surprise of many engineers. For example, a Dutch–Italian hybrid bus project resulted in exhaust emissions that were barely measurable and demonstrated very low energy consumption rates. Similar results in other experimental areas have been possible because of sophisticated combinations of small engines, flywheel systems with continuously variable transmissions, and even engine concepts that were considered obsolete, such as Sterling engines, micro gas turbines, and two-stroke engines. By now, the field of possible alternatives is so vast that several classification schemes have been proposed. The most common classification is that which distinguishes between "series hybrids," where the electric element is positioned between the thermal element and the drive wheels, and "parallel hybrids," where both the thermal and the electric element can be used separately to propel the vehicle. At the beginning of the twenty-first century, the "mild hybrid" was the latest development, in which the electric system is so small that it resembles the electric starter motor. If this development materializes, automotive history will have come full circle, producing a true compromise of an electrified gasoline car.

See also **Automobiles; Automobiles, Electric**

GIJS MOM

Further Reading

Dietmar, A. *Die Erklärung der Technikgenese des Elektroautomobils*. Frankfurt am Main, 1998.

Kirsch, D.A. Technological hybrids and the automobile system: historical considerations of the "electrified gas car" and the electrification of the automobile, in *Electric Vehicles: Socio-economic Prospects and Technological Challenges*, Cowan, R. and Hultén, S., Eds. Aldershot, 2000, pp. 74–100.

Mom, G. *History of Tomorrow's Car: Culture and Technology of the Electric Automobile*. Baltimore, forthcoming.

Mom, G. and Vincent van der Vinne, V. *De elektro-auto: een paard van Troje?* Deventer, 1995.

Sperling, D. (with contributions from Delucchi, M.A., Davis, P.M. and Burke, A.F.) *Future Drive: Electric Vehicles and Sustainable Transportation*. Covelo, CA, 1995.

Wakefield, E.H. *History of the Electric Automobile: Hybrid Electric Vehicles*. Society of Automotive Engineers, Warrendale, 1998.

Automobiles, Internal Combustion

Modern standard automobiles are driven by engines working according to the principle of internal combustion; that is, the pressure resulting from the combustion of an air and fuel mixture that is directly transformed into mechanical work. Otto engines, first built in 1877 by the German engineer Nikolaus Otto, are the standard engines for gasoline-powered cars. The only other significant and commercially successful internal combustion engine has been the diesel, developed by the German engineer Rudolf Diesel. These engines have become the standard because they are smaller, lighter, and more efficient when compared with other combustion engines of the same power.

The Otto engine operates in a continuous repetitive cycle: (1) intake of the fuel and air mixture; (2) compression, in which pressure and temperature increase slowly; (3) ignition and combustion of fuel; pressure and temperature increase rapidly; the expansion of the combustion gases forces the piston to move; pressure and temperature drop and thermal energy is converted into mechanical energy; and (4) exhaust, in which remaining pressure, or thermal energy, is released from the engine with the exhaust gases.

Today's automobile power units are highly perfected Otto and diesel engines that provide a long service life, a favorable power to mass ratio, and good start and control characteristics. If operated with conventional fuel, these cars can be driven in the range 500 to 700 kilometers on a tank of gas.

In the twentieth century, the general development of automobiles with internal combustion engines was influenced by a complex interaction between fundamental technical goals, the given technical potential, and the economic, political, and social conditions. Particularly in the early years, engine design was characterized by a host of different constructive solutions to various problems, making it difficult to detect an underlying common path of innovation. Furthermore, progress was based not only on the development of the engines themselves but also on the development of ancillary compo-

nents such as the materials, the production technology, and the fuels and lubricants.

In the late nineteenth century, there were a number of basic innovations in automobile engines. The years between the turn of the century and World War I saw a consolidation of engine design. Front mounted, water-cooled, longitudinal in-line four-cylinder four-stroke internal combustion engines became the international standard. The main goal in the early years of the century was to increase engine power, at first primarily by increasing the cubic capacity. The zenith of this tendency was reached with the famous ''Lightning Benz'' of 1909, which achieved 200 horsepower (148 kW) from a displacement of 21.5 liters of fuel. The poor efficiency, the low power density, and the difficulty of handling the inertial forces resulting from massive pistons and rods made these gigantic engines a technical dead end.

Since the end of the first decade of the twentieth century, a higher speed of rotation therefore became the major path to more powerful engines. Prerequisites for this strategy were improved lubrication (forced-feed lubrication instead of splash lubrication), a more effective charge-changing process (mechanically operated instead of vacuum operated inlet valves), and especially more efficient ignition and carburetion (high- instead of low-voltage ignition and multijet carburetors). The standard automobile engine on the eve of World War I already reached a maximum rotation speed of 2,000 revolutions per minute (rpm) and a power density of about 15 hp/l (or 11 kW/l). In comparison, Maybach's and Daimler's first automobile engine of 1885 had a maximum rotation speed of 900 rpm and 1.5 hp/l (1.1 kW/l); the average automobile engine of 1905 had a maximum 1,200 rpm and 7 hp/l (5 kW/l). Due to the limited antiknock quality of the fuel used at that time, the compression ratio of the common engines remained low, approximately 4:1.

World War I interrupted automobile but not at all engine development. To the contrary, the enormous efforts spent on the development of airplane engines led to great strides forward. New designs and materials (light alloys, nickel–steel, and chromium–nickel steel) influenced the postwar automobile engines that, in general, became more powerful and durable but nevertheless smaller and lighter. New leaded fuels allowed higher compression ratios. Until the early 1930s, the maximum rotation speed of standard engines rose to approximately 2,900 rpm, the compression ratio to about 6:1, and, as a result, the power density up to 20 hp/l (15 kW/l).

In the period between the wars, different standard designs of automobile engines emerged in Europe and the U.S. In Europe, small, fast running, four-cylinder engines prevailed due to high fuel costs and displacement dependent taxation. In the U.S., manufacturers and customers preferred huge, smooth running, high torque (but not at all fuel-efficient) six- or eight-cylinder models. American mass production technology guaranteed low production costs but, on the other hand, forced the manufacturers to produce engines principally unchanged for as long as up to three decades.

In general, the 1930s were not a time of radical change but of continuous improvement. Engines became more reliable, the average power density rose to between 25 and 30 hp/l (18 and 22 kW/l), and the engine speed increased to a maximum 4,000 rpm. The 1930s also saw the introduction of the diesel engine for passenger cars. The Daimler–Benz Company produced the first standard car with a diesel engine in 1936, and Citroen followed in 1937–1938. In those days the use of the noisy, low torque, and expensive compression ignition engines was limited to taxis.

World War II interrupted automobile development again, and in the U.S., postwar car engines were basically the same as before the war. In the 1950s however, popular demand for more power was met by further increasing the size of the engine, the compression ratio, and the rotation speed, even if these changes were at the expense of fuel economy. At the end of the 1950s, the most powerful standard engines of average medium-sized U.S. cars reached 240 hp (175 kW) and 4,500 rpm.

In Europe, postwar engine development also followed more or less familiar lines, but the manufacturers did show a higher innovative potential in details. Consequently, lightweight design (light alloy cylinder blocks, heads, and pistons) and improved carburetion and air-to-gas intake (fuel injection systems standard since 1955—bigger valves and more lift) made the still rather small engines more powerful. In the 1960s, high-performance engines with a compression ratio of about 8:1, a power density of almost 50 hp/l (37 kW), and a rotation speed of more than 5,000 rpm were found in the standard everyday car.

Since the 1960s and especially since the energy crisis of 1973 and 1974, regulatory requirements and changing consumer priorities exerted a growing influence on engine design, making emission control and fuel economy the decisive objectives of

further development. What followed was a period of rapid change primarily characterized by the introduction of microelectronic components in engine design. At first, new transistorized ignition systems allowed a fuel-saving firing. In the early 1980s, combined electronic ignition and injection systems improved engine control and, by more precise carburetion, paved the way for catalytic exhaust treatment. Today, digital engine control systems make it possible to mechanically uncouple the various components of a car engine and establish a cylinder-specific control of carburetion and ignition. Highly efficient direct fuel injection is regarded as the future of the internal combustion Otto engine in automobiles.

The 1970s energy crisis also led to an increased market share for the diesel engine automobile. Its running characteristics, performance, and fuel economy were improved, primarily by electronic control of the injection process. Furthermore, turbocharging became a common feature of standard cars with modern diesel engines. Other alternative engines with internal combustion have played a very limited role within the automobile industry. In the 1960s particularly, the Wankel engine was seen as a promising development due to specific advantages such as immediate generation of a rotary motion and comparatively simple structure. Its poor fuel economy, problematic emission levels, and the high investment requirements that a shift of production toward the rotary engine would have caused have combined to prevent its broader introduction.

Up until the early 1980s, the automobile gas turbine also seemed to be a possible alternative. However, because of further development of the conventional reciprocating engine, the principle advantages of the gas turbine (low-level raw exhaust emissions and good fuel economy at the operating point) lost relevance, whereas its specific disadvantages (poor fuel economy at part-throttle operation and high production costs) could not be overcome. Again, the high investment costs of converting mass production plants for gas turbines significantly hindered the broader introduction of this technology.

Because of their specific advantages, their high development, and their well-established production technology, gasoline and diesel internal combustion engines seemed likely to remain, at least in the medium term, the prevalent power plants of the standard car.

See also **Automobiles; Automobiles, Electric; Automobiles, Hybrid, Internal Combustion Piston Engine; Turbines, Gas in Land Vehicles**

REINHOLD BAUER

Further Reading

Caton, J.A. *History of Engine Research and Development*. American Society of Mechanical Engineers, New York, 1996.

Cummins, C.L. *Internal Fire: The Internal Combustion Engine*. Society of Automotive Engineers, Warrendale, 1989.

Cummins, C.L. *Diesel's Engine*. Carnot Press, Wilsonville, 1993.

King, D. *Computerized Engine Controls*. Delmar, Albany, 2001.

Mondt, J.R. *Cleaner Cars, The History and Technology of Emission Control Since the 1960s*. Society of Automotive Engineers, Warrendale, 2000.

Mowery, D.C. and Rosenberg, N. *Paths of Innovation: Technological Change in 20th-Century America*. Cambridge University Press, Cambridge, 1999.

Newcomb, T.P. and Spurr, R.T. *A Technical History of the Motor Car*. Adam Hilger, Bristol, 1989.

Somerscales, E.F.C. and Zagotta, A.A., Eds. *History of the Internal Combustion Engine*. American Society of Mechanical Engineers, New York, 1989.

B

Batteries, Primary and Secondary

The battery is a device that converts chemical energy into electrical energy and generally consists of two or more connected cells. A cell consists of two electrodes, one positive and one negative, and an electrolyte that works chemically on the electrodes by functioning as a conductor transferring electrons between the electrodes.

Primary cells, most often "dry cells," are exhausted (i.e., one or both of the electrodes are consumed) when they convert the chemical energy into electrical energy. These battery types are widely used in flashlights and similar devices. They generally contain carbon and zinc electrodes and an electrolyte solution of ammonium chloride and zinc chloride. Another form of primary cell, often called the mercury battery, has zinc and mercuric oxide electrodes and an electrolyte of potassium hydroxide. The mercury battery is suitable for use in electronic wristwatches and similar devices.

In 1800, Italian physicist Alessandro Volta began a new era of electrical experimentation with the voltaic pile and his discovery of the means for generating a continuous flow of electricity by chemical action for dissimilar metals separated by an electrolyte. In the nineteenth century, many different primary cells were developed and tested based on Volta's discoveries. The growth of the telegraph and railroad industries pushed battery development with their needs for various batteries with specific characteristics (e.g., sustained operation over long periods and under a wide range of temperatures). The standard form of the primary cell in the twentieth century is quite similar to the cell that Georges Leclanché invented in France in the 1860s. Another Frenchman, Felix de Lalande, also developed cells, and Thomas A. Edison improved them in the U.S. The use of batteries for electric vehicles particularly spurred Edison's interest in battery development. By 1900, the Edison–Lalande battery had displaced most of the earlier batteries, such as the Leclanché, in telegraph, telephone, and railroad signaling. Eventually, a more foolproof and maintenance-free battery, which resembled the Edison–Lalande one, displaced it in terms of number of users. In the twentieth century, household, nontechnical users of batteries rapidly outgrew the number of other users, and they preferred less maintenance to higher current strengths available in the original battery. The personal electronics revolution in the later part of the twentieth century would spur further improvements in current strengths and capacity in the maintenance-free batteries.

Secondary cells convert chemical energy into electrical energy through a chemical reaction that is essentially reversible. In "charging," the cell is forced to operate in reverse of its discharging operation by pushing a current through in the opposite direction of the one normal in discharge. Energy is thus "stored" in these cells as chemical, not electrical, energy. They may be "recharged" by an electrical current passing through them in the opposite direction of their discharge. Secondary, or storage, cells are generally wet cells, which use a liquid electrolyte.

The lead–acid storage battery and the nickel–iron, nickel–cadmium or alkaline batteries were widely used in the twentieth century. The lead–acid storage battery, mainly used in motor vehicles, consists of alternate plates of lead and lead coated with lead oxide, with an electrolyte of sulfuric acid.

The nickel–iron ("Edison") battery has nickel-plated steel grids containing tubes or pockets to hold the active materials of nickel and iron oxides and with an electrolyte of potassium hydroxide. While industry used the nickel–iron battery widely (particularly for emergency back-ups), the similar nickel-cadmium battery became the most widely used household rechargeable battery.

The development of the storage battery began with the work of French physicist Raymond Gaston Planté, who discovered the principle of the lead–acid rechargeable battery around 1859. The lead–acid battery, the most popular storage battery, is made of materials that are relatively cheap and easily manufactured, although heavy. Every alternative either requires materials that are more expensive or requires more complicated and more expensive manufacturing techniques. The storage battery industry bloomed with the advent of improved electrical generators for rapid recharging in the late nineteenth century. Until 1880, the battery was generally an apparatus for use in the laboratory because of high labor and expense for preparing and charging plates. Edison developed the nickel–iron or alkaline battery around the turn of the twentieth century. The overall structure of the storage batteries did not change much in the twentieth century. Development focused on reducing maintenance needs, reducing weight, or reducing manufacturing cost. For example, in the 1930s, battery companies began to use a new type of lead oxide produced by the companies themselves rather than provided by companies in the lead smelting, refining, and oxide-manufacturing business.

While the battery industry started in the nineteenth century with several small firms and manufacturers, by 1900 a single, large firm emerged in each of the main countries of battery development: the U.S., the U.K., Germany, and France. Companies in each country enjoyed virtual monopolies in the early part of the twentieth century as the technology of the storage battery was essentially standardized by 1920. Small differences that existed between batteries and countries depended mainly on the costs of alternate materials of construction.

Major developments in the battery industry in the twentieth century were the growing number of batteries in use by nontechnical people and the more portable structure of batteries. The automotive industry was the main push for battery development in the first half of the twentieth century. The growth of the automobile and its use of a battery for self-starting and lighting brought many people into contact with batteries. The transportation industry also required batteries for electric vehicles. Storage batteries were also widely used for emergency backups because of their relatively high capacity as a source of reserve power. New batteries developed had three basic requirements at mid-century: maximum electrical capacity per unit volume, maximum storage life under varying temperatures, and constant discharge voltage. In the second half of the twentieth century, a growing number of small, complex devices led to a great need for more compact, portable batteries. Hearing aids, pacemakers, electronic wristwatches, satellites, and personal entertainment devices spurred the number of batteries manufactured. Efforts at finding a practical battery for electric vehicles continued throughout the twentieth century. Each different need produced changes in the materials used and thus the characteristics, but not the overall design and structure. Battery development in the late twentieth century focused on changes in the electrochemical system that would operate most effectively for new and particular uses.

See also **Automobiles, Electric; Automobiles, Hybrid; Fuel Cells**

LINDA EIKMEIER ENDERSBY

Further Reading

Ruben, S. *The Evolution of Electric Batteries in Response to Industrial Needs*. Dorrance & Company, Philadelphia, 1978.

Schallenberg, R.H. *Bottled Energy: Electrical Engineering and the Evolution of Chemical Energy Storage*. American Philosophical Society, Philadelphia, 1982.

Vinal, G.W. *Primary Batteries*. Wiley, New York, 1950.

Vinal, G.W. *Storage Batteries: A General Treatise on the Physics and Chemistry of Secondary Batteries and Their Engineering Applications*, 4th edn. Wiley, New York, 1955.

Battleships

The modern battleship dates back to the final decade of the nineteenth century when the term came into general use in English for the most powerfully armed and armored surface warships. Material improvements allowed the construction of ships with high freeboard and good sea keeping capable of effectively fighting similar ships at sea, like the line of battleships of the sailing era. British battleships were the archetypes of the era. They displaced around 13,000 to 15,000 tons and their most useful armament was a battery of six 6-inch quick-firing guns on each side. These stood the best chance of successful hitting given the primitive fire

control techniques of the day, although skilled gunnery officers might use them to gain the range for accurate shooting by the slow-firing 12-inch guns, two of which were mounted in covered barbette turrets (armored structures to protect the guns) at each end.

Increasing torpedo range forced fire control improvements that emphasized longer-ranged fire with increasingly rapid-firing big guns, and the result was the entirely big gun battleship pioneered by *HMS Dreadnought* completed in 1906. Displacing 18,000 tons and armed with ten 12-inch guns in five twin turrets, she was powered by steam turbines that gave a maximum speed of 21 knots compared to 18 in the earlier ships. She made all existing pre-*Dreadnought* obsolete, and subsequent battleships were commonly known as "dreadnoughts."

Parallel with the battleship, a new type of armored cruiser had been developed in the 1890s similar in displacement to the battleship, although longer and thinner. Their larger batteries of quick-firing guns and higher speed of 23 knots made them arguably more powerful than contemporary battleships. Admiral Fisher, the dynamic British First Sea Lord usually associated with *HMS Dreadnought*, thought the all big gun armored cruiser the better type and the next ships after *Dreadnought* were the *Invincible* class armored cruisers, soon rated battle cruisers. These ships had 6-inch belt armor compared to the 11-inch maximum in *Dreadnought* but at expected combat ranges, armor gave little protection from 12-inch guns and the higher speed of 25.5 knots gave considerable tactical advantages.

Rather than replacing battleships however, battle cruisers were built alongside battleships in both the Royal Navy and its emerging rival, the Imperial German Navy. British vessels were superior in size of armament; 13.5-inch guns were adopted in the 'super dreadnoughts' commissioned in 1912 and 15-inch guns in the battleships laid down the same year. German ships were more lightly armed with 11-inch and 12-inch guns but were more heavily protected which stood them in good stead in combat. It now seems the decisive weakness in British ships was defective ammunition handling rather than protection, but the catastrophic loss of three British battle cruisers at the battle of Jutland seemed to confirm the need for armor as well as speed, especially as still longer ranges made protection more useful. New battle cruisers like *HMS Hood* were huge ships of over 40,000 tons that allowed speed to be combined with protection.

After World War I the term capital ship came into use to cover both battleships and battle cruisers. The U.S. was outbuilding Britain both in quantity and quality, and Japan also had a capital ship program that threatened both. A 5:5:3 ratio of capital ships was agreed at the Washington Conference in 1922 as was a capital ship building "holiday." The only exception to the latter was Britain's allowance of two slow 16-inch gun battleships built to a tonnage limitation of 35,000 tons "standard" displacement. There was no need for speed as the battle cruisers of this caliber armament planned by Japan and the U.S. were canceled or converted into aircraft carriers.

The Washington system broke down in the mid-1930s. Attempts to maintain qualitative restrictions limited to 14-inch guns the first class of new British capital ship, five of which were laid down as soon as was possible in 1937. The U.S.'s new capital ships mounted 16-inch guns. Japan began building four monsters of over 62,000 tons armed with nine 18-inch guns. Two, *Yamato* and *Musashi,* were completed in 1941–1942. Germany, Italy, and France, not bound by the Washington system, laid down new vessels armed with 15-inch guns. All these ships were quite fast (27–30 knots) but because of their levels of protection were known as "battleships." The term "battle cruiser" was used unofficially for German and French types of more lightly gunned battleships, respectively developments of and answers for the heavily gunned cruisers known as "pocket battleships" built by the Germans under the restrictions of the Versailles peace treaty. The fastest battleships of the new generation were the American *Iowa* class laid down in 1940–1941 which combined an armament of nine 16-inch guns, and heavy armor of up to 19.7 inches with a speed of 32.5 knots and a standard displacement of over 48,000 tons.

The speed of the *Iowas* was to allow them to operate with fast aircraft carriers. The aircraft was becoming a major challenge to the battleship. In 1940–1941 battleships were sunk at their moorings at Taranto and Pearl Harbor. Then, on 10 December 1941, the British battleship *Prince of Wales* and battle cruiser *Repulse* were sunk at sea by Japanese land-based torpedo bombers. Both *Yamato* and *Musashi* would later succumb to torpedo bombers at sea in 1944–1945, and the Italian battleship *Roma* was sunk by a German guided bomb in 1943.

Yet vulnerability was not the battleship's major problem. Sometimes the only answer to one battleship was another, as when the *Scharnhorst* was sunk by the *Duke of York* in the Arctic

darkness of December 1943. Battleships also added significantly to the anti-aircraft protection of carrier forces. Battleships were however, hungry for manpower at the same time as their supremacy as the main fleet striking unit was challenged and they thus became uneconomical. Britain and France commissioned battleships after World War II and the USSR planned abortive new units, but only the *Iowa* class survived the 1950s. All four were recommissioned in the 1980s, their armament enhanced by long-range cruise missiles. They were decommissioned in 1991–1992 but two remained in reserve ten years later. The battleship era was still not quite over.

See also **Aircraft Carriers; Radar, Defensive Systems in World War II; Ships, Bulk Carriers and Tankers; Warplanes, Fighters and Fighter Bombers**

ERIC J. GROVE

Further Reading

Breyer, S. *Battleships and Battlecruisers*. Macdonald & Janes, London, 1973.

Burt, R.A. *British Battleships 1889–1904*. Arms & Armour Press, London, 1988.

Burt, R.A. *British Battleships of World War One*. Arms & Armour Press, London, 1986.

Campbell, N.J.M. *Battlecruisers*. Conway Maritime Press, London, 1978.

Conway's *All the Worlds Fighting Ships 1860–1905, 1906–21 and 1922–46*. Conway Maritime Press, London, 1979, 1985, and 1980.

Dulin, R.O. and Garzke, W.H. *Battleships*, 3 vols. Naval Institute Press, Annapolis, 1976, 1980, and 1985.

Friedman, N. *Battleship Design and Development 1905–1945*. Conway Maritime Press, London, 1978.

Friedman, N. *US Battleships, An Illustrated Design History*. Naval Institute Press, Annapolis, 1985.

Johnston, I. and McAuley, R. *The Battleships*. Channel 4 Books, London, 2000.

Raven A. and Roberts, J. *British Battleships of World War Two*. Arms & Armour Press, London, 1976.

Bicycles and Tricycles, *see* **Transport, Human Power**

Biomass Power Generation

Biomass, or biofuels, are essentially clean fuels in that they contain no sulfur and the burning of them does not increase the long-term carbon dioxide (CO_2) levels in the atmosphere, since they are the product of recent photosynthesis (note that peat is not a biofuel in this sense). This is by no means an unimportant attribute when seen in the context of the growing awareness across the globe of the pollution and environmental problems caused by current energy production methods, and the demand for renewable energy technologies.

Biomass can be used to provide heat, make fuels, and generate electricity. The major sources of biomass include:

- Standing forests
- Wood-bark and logging residues
- Crop residues
- Short rotation coppice timber or plants
- Wood-bark mill residues
- Manures from confined livestock
- Agricultural process residues
- Seaweed
- Freshwater weed
- Algae

A few facts and figures might help to put the land-based biomass sources in perspective. The first three of the above list produce in the U.S. approximately the equivalent of 4 million barrels of oil per day in usable form. If all crop residues were collected and utilized to the full, almost 10 percent of the total U.S. energy consumption could be provided for. Although the other land-based sources of biomass are perhaps not on the same scale as this, the combined resource represents a huge untapped reservoir of potential energy. An interesting point to note is that current practices in forestry and food crop farming are aimed directly at optimizing the production of specific parts of a plant. Since biomass used for energy would make use of the whole plant, some significant advantage might be gained by growing specifically adapted crops designed to maximize the energy yield rate. It is from this origin that the energy farm concept is born.

In addition to land-based biomass, there is potential in aquatic biomass, and there are various methods by which to approach this resource. The first is by direct farming of methane as a byproduct of photosynthesis in marine plants. An example of this would be to farm huge cultivated kelp beds at great depth off the coast in suitable sea areas.

It should be noted that of the solar energy incident on the earth's surface, only 0.1 percent is harnessed through photosynthesis. Since about 2×10^{12} tons of vegetable matter grows worldwide each year, it would require only a small increase in the percentage of solar energy used in plant processes to yield a large increase in potential biomass fuel. The conversion of human and animal waste product to useful fuels has long been an interesting prospect. One method for doing just that is to employ microbial processes. It has been demonstrated that a practical regenerative system

can be developed in which waste materials are used as the feedstock upon which to grow algae. Methane can be produced by fermenting the algae and the remaining nutrient-rich waste can be recycled to grow further algae.

In a biomass farm, which combines elements of both the above techniques, algae is grown in open ponds of water in the presence of carbon dioxide and recycled inorganic nutrients. Gas lift pumps introduce CO_2 to the system and growing algae. After an incubation period, algae are collected in a trough by clumping, sedimentation, or floatation techniques, and then dewatered. The harvested biomass is deposited in a biophotolytic reactor where, under a carefully controlled environment, the algae cells use sunlight to split water molecules, forming hydrogen and oxygen.

The processes for producing energy from biomass can be divided into four areas:

1. Digestion of vegetable matter
2. Thermal processing
3. Combustion of biofuels
4. Anaerobic digestion of animal waste

The first, digestion of vegetable materials, has a resource size of 3 to 4 million tons of coal equivalent per year (Mtce per year). The economics are critically sensitive both on costs of collection and the digester equipment. The resource is characterized also by the seasonal nature of the raw material. Vegetable matter can also be converted directly into liquid fuels for transportation. The two most common biofuels are ethanol and biodiesel. Ethanol, an alcohol, is made by fermenting any biomass high in carbohydrates such as corn. It is mostly used as a fuel additive to cut down a vehicle's carbon monoxide and other smog-causing emissions. Biodiesel, an ester, is made using vegetable oils, animal fats, or algae. It can be used as a diesel additive or as a pure fuel for automobiles.

The second area, possibly more promising than the former, goes under the general heading of thermal processing. This includes the gasification, the direct liquefaction, and the pyrolysis (thermal decomposition in the absence of oxygen) of low moisture content biomass. By these means, about 5 Mtce per year of methanol alone could be produced in the U.K., 10 to 15 percent of current U.K. gasoline annual energy requirements. It is unlikely that the resource will become economically viable in the short term or even medium term.

The last two technologies are significantly better prospects and indeed have been demonstrated as commercially viable even at current fuel prices. Combustion of biofuels is said to represent a significant potential energy resource, and with most schemes having a payback period of three to five years, it presents itself as a most inviting investment. The second of these more attractive technologies is anaerobic digestion of animal wastes. The size of this resource, although still significant, is not on the same scale as the combustion of biofuels, being on the order of 1 Mtce of economically viable potential. Although much work on this resource has been carried out among farming cooperatives in Denmark, there are a number of uncertainties that significantly affect the rate of development of this resource.

The product of this process is methane, and it is most likely that the exploitation of the resource will be carried out on a local scale, perhaps at farm level. Thus the marketability of surplus methane; that is, that which is in excess of the farmer's needs, is not certain. Although the technology for digester construction is well established, the actual processes that occur during operation are still poorly understood. Consequently there are uncertainties as to the design performance and flexibility in adapting to any fuel variations that may occur.

Despite these drawbacks and the general unsuitability of high technology to the agricultural environment, research is continuing along a number of promising lines that could lead to increased controllability and the reduction of costs on less economic fuels such as dairy cattle waste. Although the manufacture of digesters for the more attractive fuels such as pig and poultry waste is well established, it remains to be seen how quickly the technology will be taken up and perform in a working environment.

See also **Electricity Generation and the Environment; Energy and Power; Power Generation, Recycling; Solar Power Generation; Wind Power Generation**

IAN BURDON

Further Reading

Andreae, M.O. Biomass burning: its history, use, and distribution and its impact on environmental quality and global climate, in *Global Biomass Burning: Atmospheric, Climatic, and Biospheric Implications*, J.S. Levine, Ed. MIT Press, Cambridge, MA, 1991, pp. 1–21.

Goldemberg, J. *Energy, Environment and Development*, Earthscan, London, 1996.

Kolar, J. Renewable energy: biomass power and biofuels as alternative sustainable energy sources. *Environ. Qual. Manag.*, 8, 2, 35–42, 1998.

Leggett, J. *The Carbon War. Global Warming at the End of the Oil Era*. Penguin, London, 1999.

MacNeill, J.R., Ed. *Something New Under the Sun: An Environmental History of the Twentieth Century.* Penguin, London, 2001.

Biopolymers

Biopolymers are natural polymers, long-chained molecules (macromolecules) consisting mostly of a repeated composition of building blocks or monomers that are formed and utilized by living organisms. Each group of biopolymers is composed of different building blocks, for example chains of sugar molecules form starch (a polysaccharide), chains of amino acids form proteins and peptides, and chains of nucleic acid form DNA and RNA (polynucleotides). Biopolymers can form gels, fibers, coatings, and films depending on the specific polymer, and serve a variety of critical functions for cells and organisms. Proteins including collagens, keratins, silks, tubulins, and actin usually form structural composites or scaffolding, or protective materials in biological systems (e.g., spider silk). Polysaccharides function in molecular recognition at cell membrane surfaces, form capsular barrier layers around cells, act as emulsifiers and adhesives, and serve as skeletal or architectural materials in plants. In many cases these polymers occur in combination with proteins to form novel composite structures such as invertebrate exoskeletons or microbial cell walls, or with lignin in the case of plant cell walls.

Natural biopolymers have a huge diversity in functions, yet only a small number of this diverse group of polymers has been extensively studied and commercially exploited. However, the impact of even these limited numbers in the twentieth century has been substantive. Biopolymers from plants (e.g., starch, cellulose in cotton and flax, natural rubber) and animals (e.g., collagen or gelatin) have traditionally been gathered and utilized for centuries as food and materials. In the twentieth century, additional biopolymers have been processed or extracted from plants (e.g., alginates from seaweeds) and exploited in many important materials-related applications, and new polymers have been designed or engineered (see below). Alginate from seaweed was developed in the 1930s when the Kelco Company commercialized the biopolymer as a food stabilizer in ice cream. Biopolymers have since transformed food industries as gums, stabilizers, and emulsifiers, the oil extraction industry in the use of xanthan to enhance oil recovery, medicine (e.g., collagen and silk biomaterials), and cellulose and starch used extensively in the pulp, paper, and textile industries. The rheological properties of biopolymers such as xanthan give them their useful properties: adding xanthan to water-based drilling fluids enhances the viscosity.

Xanthan is also used in food and pharmaceutical formulations due to its stability over a broad range of pH and salt concentrations. Alginates are used for gelation, emulsification, and stabilization in foods and ceramic formulations, as coatings for paper, and in pharmaceutical formations. Biodegradable starch-based products and bacterial polyhydroxyalkanoates generated by thermal extrusion or solution processing have been extensively studied in recent years as lower cost options for commodity plastics and for biomedical material applications.

Traditional modes of generating biopolymers have included farming, as in the case of cellulose from cotton and starch from corn. New methods of generating polymeric materials from plants may involve:

1. Improved modes of production via tissue culturing in bioreactors
2. Degrading the plant material chemically or by microbial fermentation and subsequently synthesizing new polymers
3. Tailoring the original polymeric structure of the plant material by enzymatic synthesis or genetic engineering. A variety of microbial polysaccharides produced outside of cells are produced commercially; probably the most important is xanthan gum, produced by enzymes from corn syrup. Advances in the *in vitro* synthesis of biopolymers, via cell culture systems or enzymatic catalysis, suggest future opportunities to further expand on the suite of monomers amenable to direct incorporation into biopolymers during biological synthesis. New options are being actively explored to expand the building blocks (monomers such as amino acids, sugars and fatty acids) used in these types of polymers by marrying chemistry and biology. The incorporation of nonnative building blocks, such as modified amino acids into proteins, alternative sugars into polysaccharides, and alternative fatty acids into polyesters, are examples of the expanding range of monomers that can be utilized in biopolymers that can be generated by biological synthesis, an approach that avoids the historical limitations imposed by biology and evolution or selection. For example, fluorinated amino acids and fluorinated fatty

> acids have been successfully incorporated into proteins and into polysaccharides like emulsan, a biopolymer synthesized naturally by a bacterium, and which is amenable to structural tailoring.

Extensive chemical derivatization of polysaccharides is carried out industrially, particularly with starches (mostly derived from corn) and cellulose, as a means to alter solubility and properties. Genetic engineering is being pursued via transgenic animals or plants, by inserting bacterial genes that create new biosynthetic pathways. For example, transgenic collagens and silks are being pursued as a route to more cost-effective sources of these materials for biomedical and commodity materials. Genes isolated from spider species could be used in mammalian cells to produce silk.

Polyhydroxylalkanoates (PHAs) are a large family of structurally related polyesters synthesized by many bacteria or transgenic plants and accumulated as granules within the cell. These natural thermoplastics were investigated in the 1970s following the oil crisis, and developed by ICI in the U.K. These polymers are composed of intracellular homo- or copolymers of [R]-β-hydroxyalkanoic acids. PHAs have been pursued for biomedical materials as well as replacements for petrochemically derived polymers since they biodegrade naturally and completely, and have diverse options for monomer chemistries that can be incorporated biologically into the polymer. Once extracted from the cells and processed into plastics, some PHAs exhibit material properties similar to polypropylene. PHAs have been produced on an industrial scale (one polymer produced under the trade name Biopol is used as a packaging material, for example for shampoo bottles), although commercial applications for PHAs have focused in recent years on biomedical materials applications due to the high costs of bacterially synthesized polymer production.

An important feature of biological synthesis is the template-directed process used in the case of proteins and nucleic acid biosynthesis. In comparison to synthetic approaches to polymer synthesis, this method provides direct control over monomer sequence (and thus chemistry) and size of the polymers. Therefore, biosynthesis methods are carefully orchestrated processes aimed at optimizing structures for molecular recognition to drive macromolecular assembly, while also designed to conserve resources to promote survival of the organism. These polymers are also recycled back into natural geochemical and biological cycles for reuse and serve as models for "green chemistry" approaches.

Advances in metabolic engineering, environmental considerations about renewable polymers from nonpetroleum feedstocks, and the expansion in molecular biology and protein engineering tools in general are taking biopolymer synthesis and production in new directions. The opportunity to enhance, alter, or direct the structural features of biopolymers through genetic manipulation, physiological controls, or enzymatic processes provides new routes to novel polymers with specialty functions. The use of biopolymers in commodity and specialty materials, as well as biomedical applications, can be expected to continue to increase with respect to petrochemical-derived materials. The benefit in tailoring structural features is a plus for generating higher performance properties or more specialized functional performance. Biosynthesis and disposal of biopolymers can be considered within a renewable resource loop, reducing environmental burdens associated with synthetic polymers derived from petrochemicals that often require hundreds of years to degrade. In addition, biopolymers can often be produced from low cost agricultural feedstocks versus petroleum supplies and thereby generate value-added products.

See also **Biotechnology; Synthetic Resins**

DAVID KAPLAN

Further Reading

Byrom, D. Polymer synthesis by micro-organisms: technology and economics. *Trends Biotechnol.* 5, 246–250, 1987.

Kaplan, D.L. Introduction to Biopolymers from Renewable Resources, in *Biopolymers from Renewable Resources*, Kaplan, D.L., Ed. Springer Verlag, Berlin, 1998, pp. 1–26.

Shogrun, R.L. 1998. Starch properties and material applications, in *Biopolymers from Renewable Resources*, Kaplan, D.L., Ed. Springer Verlag, Berlin, 1998, pp. 30-46

Steinbuchel, A. and Valentin. H.E. Diversity of bacterial polyhydroxyalkanoic acids. *FEMS Microbiol. Lett.* 128, 219–228, 1995.

Biotechnology

The term biotechnology came into popular use around 1980 and was understood to mean the industrial use of microorganisms to make goods and services (Commission of the European Communities, 1979). Although biotechnology is often associated with the application of genetics, that is too narrow an interpretation. Rather the

Figure 1. Different genetic sources of alfalfa are evaluated to identify plant traits that would increase growth and enhance the conversion of plant tissues into biofuel. [*Photo by Keith Weller. ARS/USDA*].

word has been used for almost a century to reflect a changing combination of the manipulation of organisms, the means of multiplying them using fermentation, and the extraction of useful products. Moreover, while the technology of the 1980s was new, claims that the introduction of biotechnology would mark a new industrial revolution had been made with conviction and vision since the time of World War I.

Biotechnology grew out of the technology of fermentation, which was called zymotechnology. This was different from the ancient craft of brewing because of its thought-out relationships to science. These were most famously conceptualized by the Prussian chemist Georg Ernst Stahl (1659–1734) in his 1697 treatise *Zymotechnia Fundamentalis*, in which he introduced the term zymotechnology. Carl Balling, long-serving professor in Prague, the world center of brewing, drew on the work of Stahl when he published his *Bericht über die Fortschritte der zymotechnische Wissenschaften und Gewerbe* (Account of the Progress of the Zymotechnic Sciences and Arts) in the mid-nineteenth century. He used the idea of zymotechnics to compete with his German contemporary Justus Liebig for whom chemistry was the underpinning of all processes.

By the end of the nineteenth century, there were attempts to develop a new scientific study of fermentation. It was an aspect of the "second" Industrial Revolution during the period from 1870 to 1914. The emergence of the chemical industry is widely taken as emblematic of the formal research

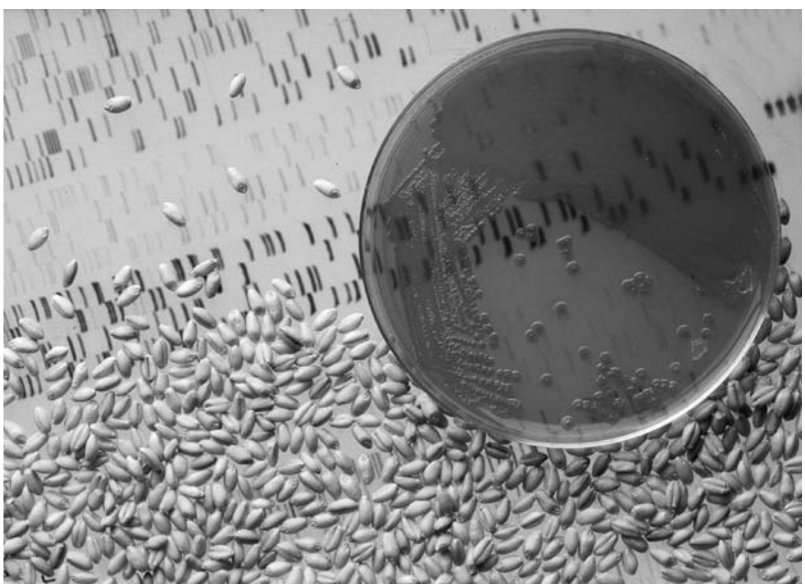


Figure 2. Wheat seeds coated with genetically modified bacteria like those colonized in this petri dish are nearly immune to wheat take-all, a root-destroying fungal disease. The sequencing gel in the background bears the genetic code for bacterial enzymes that synthesize natural antibiotics. [*Photo by Jack Dykinga. ARS/USDA*].

and development taking place at the time. The development of microbiological industries is another example. For the first time, Louis Pasteur's germ theory made it possible to provide convincing explanations of brewing and other fermentation processes.

Pasteur had published on brewing in the wake of France's humiliation in the Franco–Prussian war (1870–1871) to assert his country's superiority in an industry traditionally associated with Germany. Yet the science and technology of fermentation had a wide range of applications including the manufacture of foods (cheese, yogurt, wine, vinegar, and tea), of commodities (tobacco and leather), and of chemicals (lactic acid, citric acid, and the enzyme takaminase). The concept of zymotechnology associated principally with the brewing of beer began to appear too limited to its principal exponents. At the time, Denmark was the world leader in creating high-value agricultural produce. Cooperative farms pioneered intensive pig fattening as well as the mass production of bacon, butter, and beer. It was here that the systems of science and technology were integrated and reintegrated, conceptualized and reconceptualized.

The Dane Emil Christian Hansen discovered that infection from wild yeasts was responsible for numerous failed brews. His contemporary Alfred Jørgensen, a Copenhagen consultant closely associated with the Tuborg brewery, published a widely used textbook on zymotechnology. *Microorganisms and Fermentation* first appeared in Danish 1889 and would be translated, reedited, and reissued for the next 60 years.

The scarcity of resources on both sides during World War I brought together science and technology, further development of zymotechnology, and formulation of the concept of biotechnology. Impending and then actual war accelerated the use of fermentation technologies to make strategic materials. In Britain a variant of a process to ferment starch to make butadiene for synthetic rubber production was adapted to make acetone needed in the manufacture of explosives. The process was technically important as the first industrial sterile fermentation and was strategically important for munitions supplies. The developer, chemist Chaim Weizmann, later became well known as the first president of Israel in 1949.

In Germany scarce oil-based lubricants were replaced by glycerol made by fermentation. Animal feed was derived from yeast grown with the aid of the new synthetic ammonia in another wartime development that inspired the coining of the word biotechnology. Hungary was the agricultural base of the Austro–Hungarian empire and aspired to Danish levels of efficiency. The economist Karl Ereky (1878–1952) planned to go further and build the largest industrial pig-processing factory. He envisioned a site that would fatten 50,000 swine at a time while railroad cars of sugar beet arrived and fat, hides, and meat departed. In this forerunner of the Soviet collective farm, peasants (in any case now falling prey to the temptations of urban society) would be completely superseded by the industrialization of the biological process in large factory-like animal processing units. Ereky went further in his ruminations over the meaning of his innovation. He suggested that it presaged an industrial revolution that would follow the transformation of chemical technology. In his book entitled *Biotechnologie,* he linked specific technical injunctions to wide-ranging philosophy. Ereky was neither isolated nor obscure. He had been trained in the mainstream of reflection on the meaning of the applied sciences in Hungary, which would be remarkably productive across the sciences. After World War I, Ereky served as Hungary's minister of food in the short-lived right wing regime that succeeded the fall of the communist government of Bela Kun.

Nonetheless it was not through Ereky's direct action that his ideas seem to have spread. Rather, his book was reviewed by the influential Paul Lindner, head of botany at the Institut für Gärungsgewerbe in Berlin, who suggested that microorganisms could also be seen as biotechnological machines. This concept was already found in the production of yeast and in Weizmann's work with strategic materials, which was widely publicized at that very time. It was with this meaning that the word "Biotechnologie" entered German dictionaries in the 1920s.

Biotechnology represented more than the manipulation of existing organisms. From the beginning it was concerned with their improvement as well, and this meant the enhancement of all living creatures. Most dramatically this would include humanity itself; more mundanely it would include plants and animals of agricultural importance. The enhancement of people was called eugenics by the Victorian polymath and cousin of Charles Darwin, Francis Galton. Two strains of eugenics emerged: negative eugenics associated with weeding out the weak and positive eugenics associated with enhancing strength. In the early twentieth century, many eugenics proponents believed that the weak could be made strong. People had after all progressed beyond their biological limits by means of technology.

Jean-Jacques Virey, a follower of the French naturalist Jean-Baptiste de Monet de Lamarck, had coined the term "biotechnie" in 1828 to describe man's ability to make technology do the work of biology, but it was not till a century later that the term entered widespread use. The Scottish biologist and town planner Patrick Geddes made biotechnics popular in the English-speaking world. Geddes, too, sought to link life and technology. Before World War I he had characterized the technological evolution of mankind as a move from the paleotechnic era of coal and iron to the neotechnic era of chemicals, electricity, and steel. After the war, he detected a new era based on biology—the biotechnic era. Through his friend, writer Lewis Mumford, Geddes would have great influence. Mumford's book *Technics and Civilization,* itself a founding volume of the modern historiography of technology, promoted his vision of the Geddesian evolution.

A younger generation of English experimental biologists with a special interest in genetics, including J. B. S. Haldane, Julian Huxley, and Lancelot Hogben, also promoted a concept of biotechnology in the period between the world wars. Because they wrote popular works, they were among Britain's best-known scientists. Haldane wrote about biological invention in his far-seeing work *Daedalus*. Huxley looked forward to a blend of social and eugenics-based biological engineering. Hogben, following Geddes, was more interested in engineering plants through breeding. He tied the progressivism of biology to the advance of socialism.

The improvement of the human race, genetic manipulation of bacteria, and the development of fermentation technology were brought together by the development of penicillin during World War II. This drug was successfully extracted from the juice exuded by a strain of the *Penicillium* fungus. Although discovered by accident and then developed further for purely scientific reasons, the scarce and unstable "antibiotic" called penicillin was transformed during World War II into a powerful and widely used drug. Large networks of academic and government laboratories and pharmaceutical manufacturers in Britain and the U.S. were coordinated by agencies of the two governments. An unanticipated combination of genetics, biochemistry, chemistry, and chemical engineering skills had been required. When the natural mold was bombarded with high-frequency radiation, far more productive mutants were produced, and subsequently all the medicine was made using the product of these man-made cells. By the 1950s penicillin was cheap to produce and globally available.

The new technology of cultivating and processing large quantities of microorganisms led to calls for a new scientific discipline. Biochemical engineering was one term, and applied microbiology another. The Swedish biologist, Carl-Goran Heden, possibly influenced by German precedents, favored the term "Biotechnologi" and persuaded his friend Elmer Gaden to relabel his new journal *Biotechnology and Biochemical Engineering*. From 1962 major international conferences were held under the banner of the *Global Impact of Applied Microbiology*. During the 1960s food based on single-cell protein grown in fermenters on oil or glucose seemed, to visionary engineers and microbiologists and to major companies, to offer an immediate solution to world hunger. Tropical countries rich in biomass that could be used as raw material for fermentation were also the world's poorest. Alcohol could be manufactured by fermenting such starch or sugar rich crops as sugar cane and corn. Brazil introduced a national program of replacing oil-based petrol with alcohol in the 1970s.

It was not, however, just the developing countries that hoped to benefit. The Soviet Union developed fermentation-based protein as a major source of animal feed through the 1980s. In the U.S. it seemed that oil from surplus corn would solve the problem of low farm prices aggravated by the country's boycott of the USSR in1979, and the term "gasohol" came into currency. Above all, the decline of established industries made the discovery of a new wealth maker an urgent priority for Western governments. Policy makers in both Germany and Japan during the 1970s were driven by a sense of the inadequacy of the last generation of technologies. These were apparently maturing, and the succession was far from clear. Even if electronics or space travel offered routes to the bright industrial future, these fields seemed to be dominated by the U.S. Seeing incipient crisis, the Green, or environmental, movement promoted a technology that would depend on renewable resources and on low-energy processes that would produce biodegradable products, recycle waste, and address problems of the health and nutrition of the world.

In 1973 the German government, seeking a new and "greener" industrial policy, commissioned a report entitled *Biotechnologie* that identified ways in which biological processing was key to modern developments in technology. Even though the report was published at the time that recombinant

DNA (deoxyribonucleic acid) was becoming possible, it did not refer to this new technique and instead focused on the use and combination of existing technologies to make novel products.

Nonetheless the hitherto esoteric science of molecular biology was making considerable progress, although its practice in the early 1970s was rather distant from the world of industrial production. The phrase "genetic engineering" entered common parlance in the 1960s to describe human genetic modification. Medicine, however, put a premium on the use of proteins that were difficult to extract from people: insulin for diabetics and interferon for cancer sufferers. During the early 1970s what had been science fiction became fact as the use of DNA synthesis, restriction enzymes, and plasmids were integrated. In 1973 Stanley Cohen and Herbert Boyer successfully transferred a section of DNA from one *E. coli* bacterium to another. A few prophets such as Joshua Lederberg and Walter Gilbert argued that the new biological techniques of recombinant DNA might be ideal for making synthetic versions of expensive proteins such as insulin and interferon through their expression in bacterial cells. Small companies, such as Cetus and Genentech in California and Biogen in Cambridge, Massachusetts, were established to develop the techniques. In many cases discoveries made by small "boutique" companies were developed for the market by large, more established, pharmaceutical organizations.

Many governments were impressed by these advances in molecular genetics, which seemed to make biotechnology a potential counterpart to information technology in a third industrial revolution. These inspired hopes of industrial production of proteins identical to those produced in the human body that could be used to treat genetic diseases. There was also hope that industrially useful materials such as alcohol, plastics (biopolymers), or ready-colored fibers might be made in plants, and thus the attractions of a potentially new agricultural era might be as great as the implications for medicine. At a time of concern over low agricultural prices, such hopes were doubly welcome. Indeed the agricultural benefits sometimes overshadowed the medical implications.

The mechanism for the transfer of enthusiasm from engineering fermenters to engineering genes was the New York Stock Exchange. At the end of the 1970s, new tax laws encouraged already adventurous U.S. investors to put money into small companies whose stock value might grow faster than their profits. The brokerage firm E. F. Hutton saw the potential for the new molecular biology companies such as Biogen and Cetus. Stock market interest in companies promising to make new biological entities was spurred by the 1980 decision of the U.S. Supreme Court to permit the patenting of a new organism. The patent was awarded to the Exxon researcher Ananda Chakrabarty for an organism that metabolized hydrocarbon waste. This event signaled the commercial potential of biotechnology to business and governments around the world. By the early 1980s there were widespread hopes that the protein interferon, made with some novel organism, would provide a cure for cancer. The development of monoclonal antibody technology that grew out of the work of Georges J. F. Köhler and César Milstein in Cambridge (co-recipients with Niels K. Jerne of the Nobel Prize in medicine in 1986) seemed to offer new prospects for precise attacks on particular cells.

The fear of excessive regulatory controls encouraged business and scientific leaders to express optimistic projections about the potential of biotechnology. The early days of biotechnology were fired by hopes of medical products and high-value pharmaceuticals. Human insulin and interferon were early products, and a second generation included the anti-blood clotting agent tPA and the antianemia drug erythropoietin. Biotechnology was also used to help identify potential new drugs that might be made chemically, or synthetically.

At the same time agricultural products were also being developed. Three early products that each raised substantial problems were bacteria which inhibited the formation of frost on the leaves of strawberry plants (ice-minus bacteria), genetically modified plants including tomatoes and rapeseed, and the hormone bovine somatrotropin (BST) produced in genetically modified bacteria and administered to cattle in the U.S. to increase milk yields. By 1999 half the soy beans and one third of the corn grown in the U.S. were modified. Although the global spread of such products would arouse the best known concern at the end of the century, the use of the ice-minus bacteria—the first authorized release of a genetically engineered organism into the environment—had previously raised anxiety in the U.S. in the 1980s.

In 1997 Dolly the sheep was cloned from an adult mother in the Roslin agricultural research institute outside Edinburgh, Scotland. This work was inspired by the need to find a way of reproducing sheep engineered to express human proteins in their milk. However, the public interest was not so much in the cloning of sheep that had just been achieved as in the cloning of people,

which had not. As in the Middle Ages when deformed creatures had been seen as monsters and portents of natural disasters, Dolly was similarly seen as monster and as a portent of human cloning.

The name *Frankenstein*, recalled from the story written by Mary Shelley at the beginning of the nineteenth century and from the movies of the 1930s, was once again familiar at the end of the twentieth century. Shelley had written in the shadow of Stahl's theories. The continued appeal of this book embodies the continuity of the fears of artificial life and the anxiety over hubris. To this has been linked a more mundane suspicion of the blending of commerce and the exploitation of life. Discussion of biotechnology at the end of the twentieth century was therefore colored by questions of whose assurances of good intent and reassurance of safety could be trusted.

See also **Breeding, Animal, Genetic Methods; Breeding, Plant, Genetic Methods; Cell Culturing; Cloning, Testing and Treatment Methods; Feedstocks; Genetic Engineering, Methods; Genetic Engineering, Applications; Medicine; Synthetic Foods**

ROBERT BUD

Further Reading

Beier, F.K., Crespi, R.S. and Strauss, J. *Biotechnology and Patent Protection: An International Review*. OECD, Paris, 1985. Classic overview of patenting issues.

Bud, R. *Uses of Life: A History of Biotechnology,* Cambridge University Press, Cambridge, 1994. This volume explores many of the issues dealt with in this piece.

Commission of the European Communities. FAST: Subprogramme C 'Biosociety', FAST/ACPM, 79, 1G–3E, 1979. One of the very first overall analyses of the field, its findings were included in the book published as *Eurofutures*.

FAST Group. *Eurofutures: The Challenge of Innovation*. Butterworth, London, 1987.

Gaskell, G. and Bauer, M., Eds. *Biotechnology: The years of controversy*. NMSI Trading, London, 2001.

Bauer, M.W. and Gaskell, G., Eds. *Biotechnology: The Making of a Global Controversy*. Cambridge University Press, Cambridge, 2002.

Hall, S. *Invisible Frontiers: The Race to Synthesise a Human Gene,* 2nd edn. Oxford University Press, Oxford, 2002. This tells the story of the race between Biogen and Genentech to make human insulin one of the very first products of genetic engineering.

Kevles, D. *In the Name of Eugenics: Genetics and the Uses of Human Heredity*. Alfred A. Knopf, New York, 1985.

Kloppenberg, J.R., Jr. *First the Seed: The Political Economy of Plant Biotechnology 1492–2000*. Cambridge University Press, Cambridge, 1988.

Orsenigo, L. *The Emergence of Biotechnology: Institutions and Markets in Industrial Innovation*. Pinter, London, 1989. Overview of the birth of biotechnology.

Parascandola, J., Ed. *The History of Antibiotics: A Symposium*. American Institute of the History of Pharmacy, Madison, 1980.

Rothman, H., Greenshields, R. and Callé, F.R. *The Alcohol Economy: Fuel Ethanol and the Brazilian Experience*. Pinter, London 1983.

Teitelman R. *Gene Dreams: Wall Street, Academia, and the Rise of Biotechnology*. Basic Books, New York, 1989.

Thackray A., Ed. *Private Science: Biotechnology and the Rise of the Biomolecular Sciences*. University of Pennsylvania Press, Philadelphia, 1998.

Turney J. *Frankenstein's Footsteps: Science, Genetics and Popular Culture*. Yale University Press, New Haven, 1998.

Wilmut, I., Campbell, K. and Tudge, C. *The Second Creation: The Art of Biological Control by the Scientists Who Cloned Dolly*. Headline, London, 2000.

Blood Transfusion and Blood Products

Blood has always had a cultural significance, symbolic of the essence of life; but the process of transfusion—replacing blood with blood—only became an accepted and reliable practice in the twentieth century.

William Harvey's demonstration of blood circulation in 1628 opened up the possibility of transfusion. In 1665 an English physiologist, Richard Lower, described the first successful transfusion between dogs. The first human transfusion came two years later: Frenchman Jean-Baptiste Denis transferred blood from a lamb to a sick boy, who reportedly recovered. The experiment was repeated but, following several deaths, was banned by 1678.

Interest revived in the nineteenth century when the role of blood as an oxygen transporter was understood. James Blundell at Guy's Hospital in London used transfusions to revive women who hemorrhaged after childbirth. But there were two main problems. First, outside the body, blood would quickly clot, stopping free flow. Second, many patients had severe, sometimes fatal, reactions. Karl Landsteiner solved this second problem in 1901 with his discovery of blood groups. He noticed a human serum sample "clumped" the red blood cells from some people but not others. Using new immunological theories, Landsteiner realized agglutination was due to the presence or absence of specific antigens on the red blood cell. Some individuals have antigen A, some B, some both, and some neither, leading to four blood groups or types: A, B, AB, and O. Not all groups are compatible; mixing incompatible groups causes potentially fatal clumping. Landsteiner's discovery

would ultimately make blood transfusion safe (he was awarded the Nobel Prize for physiology in 1930 for this work), but clinicians initially ignored the importance. In 1908 Reuben Ottenberg introduced typing and cross-matching of donors and recipients, but compatibility testing was not immediately adopted.

Improvements in surgery drove the early twentieth century reintroduction of transfusion in America. In 1902 Alexis Carrel reported the possibility of direct transfusion, sewing a donor's artery to the recipient's vein (anastomosis). George Crile pioneered the technique, carrying out over 200 animal transfusions before progressing to humans, but Carrel received the recognition following publicity describing a transfusion between a surgeon and his 5-day-old son.

Direct transfusion avoided problems of coagulation but required delicate and painstaking surgery. It was difficult to quantify the amount of blood transferred, which could be lethal. Other surgeons experimented with semidirect methods of transfusion. W. G. Kimpton and M.S. Brown, along with many others, developed specialized equipment using canola, syringe, needles, stopcocks, and valves. Coating vessels with paraffin wax minimized clotting. Coagulation was overcome in 1914; three doctors (Agote in Argentina, Hustin in Belgium, and Lewisohn in the U.S.) independently demonstrated that sodium citrate could be used as an anticoagulant. Adding small concentrations to blood did not harm the patient but prevented clotting. Indirect transfusion was now possible.

World War I accelerated the pace of change. With increasing demand, the indirect method was perfected. Blood was collected in citrate–glucose solution, refrigerated and transported in bottles to the front lines. Transfusion spread from North America to previously skeptical Europe. By the war's end, it was a practical and relatively simple treatment that saved thousands of lives. The focus then turned to donor recruitment. The need for blood typing became clear; rapid testing procedures allowed selection of appropriate blood. In 1921, Percy Lane Oliver set up the first transfusion service with the British Red Cross. It was a "walking donor" service in which volunteers of known blood groups were available on demand, donating blood wherever it was needed. The idea spread, and donor panels were set up in Europe, the U.S., and the Far East during the 1920s and 1930s. The first blood bank was established in 1932 at Leningrad Hospital in Russia.

The outbreak of World War II prompted another dramatic expansion in blood donation services. A huge logistical operation supplied blood to the front lines and to civilian casualties; by 1944 U.K. donors provided 1200 pints a day.

Plasma, the yellow serum that carries red cells, became a common transfusion fluid, used to treat shock by restoring blood volume. Using Flosdorf and Mudd's lyophilization process, plasma was freeze-dried. Removing water under high vacuum left a dry powder, stable for months. Adding sterile water reconstituted the plasma. Other plasma warwork had long-term impact. Edwin Cohn, from Harvard Medical School, developed a process of cold ethanol fractionation to break plasma down into components. The most important product, albumin, was isolated from Fraction V. Packaged in glass ampoules, this concentrated ready-to-use liquid had vital antishock capabilities. Other products were developed from fractions: gamma globulin, fibrin foam, and blood-grouping globulins. The Plasma Fractionation Project expanded to an industrial scale, with collaboration between universities and pharmaceutical companies.

After the war, civilian blood transfusion expanded. The U.K. Blood Transfusion Service was established in 1946, recruiting voluntary donors with the promise of a cup of tea. More controversially, donors in the U.S. were paid. Developments in blood typing and screening ensured compatibility. Over twenty genetically determined blood group systems were identified, including Rhesus positive and negative.

Collection equipment improved as disposable equipment replaced glass flasks and rubber tubing. In 1950 Carl Walter introduced the plastic collecting bag, having experimented with polymers to find one suitably robust, inert, and immune to extreme temperatures. The new bags reduced contamination and allowed economical ultra-low temperature freezing of blood. Using cryoprotectants like glycerol, red blood cells were preserved for long periods, allowing stockpiling of rare blood types.

Processing developments continued through the century. Today, blood is collected into 450-ml plastic packs with anticoagulant solution. Using a closed system of satellite bags, it is centrifuged with minimal risk of contamination. The red cells, platelets, and plasma components are separated into individual bags, ready for further processing. More than 17 preparations of blood components are available, including clotting factors (such as Factor VIII for hemophiliacs), and antibodies for vaccine production. Whole blood is used only rarely, but no part of a blood donation is wasted.

The wide availability of blood components has facilitated dramatic advances in surgery. Blood

transfusion is commonplace in hospitals and clinical blood transfusion is a specialty in its own right. In the U.K. alone, over 2.5 million units of blood are collected each year, and demand continues to rise. In the last two decades of the century however, there was concern about virus transmission through transfusion. Public scandals in France, Canada, and Japan, where patients and particularly hemophiliacs became infected with HIV as a result of transfusions, led to comprehensive monitoring at all stages of donation. Testing for HIV was introduced in 1986 and for hepatitis C in 1991. Blood labeling was internationally standardized in 1992.

Artificial blood substitutes may be the future. Blood volume expanders and hemodilutants (isotonic electrolyte solutions) are already widely used. The search for an artificial oxygen transporter is underway. Possibilities include microencapsulated hemoglobin, recombinant hemoglobin, or perfluorochemical emulsions.

See also **Hematology**

NICOLA PERRIN

Further Reading

Wintrobe, M., Ed. *Blood, Pure and Eloquent*. McGraw-Hill, New York, 1980.

Starr, D. *Blood: An Epic History of Medicine and Commerce*. Alfred A. Knopf, 1998.

Creager, A. Biotechnology and Blood: Edwin Cohn's Plasma Fractionation Project, 1940–1953, in *Private Science: Biotechnology and the Rise of the Molecular Sciences*, Thackray, A., Ed. University of Pennsylvania Press, Philadelphia, 1998.

Contreras, M., Ed. *ABC of Transfusion*. BMJ Books, London, 1998.

Keynes, G. *Blood Transfusion*. John Wright & Son, Bristol, 1949.

Schneider, W. Blood transfusion in peace and war, 1900–1918. *Soc. Hist. Med.*, 10, 01, 1997, 105–126.

Useful websites

National Blood Service (UK): http://www.blood.co.uk

Blood Transfusion Safety, World Health Organization: http://www.who.int/bct/

Bombs, *see* **Fission and Fusion Bombs**

Boranes

Boranes are chemical compounds of boron and hydrogen. During the 1950s, the U.S. government sponsored a major secretive effort to produce rocket and aircraft fuels based on boron hydrides. Much of the information initially available to the U.S. effort was contained in a book written by the German chemist Alfred Stock in 1933. When burned in air, the energy released by various boron hydride compounds, as measured by their heat of combustion, is 20 to 55 percent greater than the energy released by petroleum-based jet fuels. It was expected that this greater energy content of the boron fuels would translate into equivalent higher payloads or ranges for rockets and aircraft.

All of the boron fuel manufacturing processes started with the production of diborane, a compound composed of two boron atoms linked to six hydrogen atoms. Initially, this was produced by reacting lithium hydride with boron trifluoride or boron trichloride in diethyl ether as a solvent. This entailed a need to recover and recycle the expensive lithium. A later process produced diborane by reacting sodium borohydride with boron trichloride in the presence of catalytic amounts of aluminum chloride, using a solvent called diglyme.

Diborane is a gas that is highly toxic and pyrophoric (can catch fire spontaneously on contact with air), and the boron hydrides burn with a brilliant green flame. Diborane was condensed to a liquid at $-80^{\circ}C$ and transferred to high-pressure cylinders which were stored at $-10^{\circ}C$ to minimize degradation in storage.

The diborane had to be converted to liquids suitable for storage in aircraft or rocket fuel tanks at the normal operating temperature for those tanks. By 1952, diborane and pentaborane were being produced in pilot plant quantities. There were two major fuel-producing contractors on the project. One of the two produced liquid fuels by alkylating the diborane with ethylene. The other contractor pyrolyzed the diborane to pentaborane and decaborane and then alkylated those with propyl, ethyl, or methyl groups.

Pyrolysis is a process in which diborane is diluted with hydrogen and circulated through a carefully heated tube in which 60 to 70 percent of the diborane converts to pentaborane, a lesser amount converts to decaborane, and the remainder converts to waste boron hydrides (which would have to be recycled to recover the relatively scarce boron) and hydrogen gas. Alkylation is a process in which hydrocarbon groups were attached to the boranes, for example by reacting pentaborane with propyl chloride in the presence of aluminum chloride catalyst to produce propyl pentaborane. Alkylation produced the desired physical properties at the expense of reducing the heat of combustion of the resulting fuel in direct proportion to the amount of hydrocarbon added.

Pentaborane is a toxic, colorless, pyrophoric room temperature liquid with a 60°C boiling point. There are two forms, but the stable compound consists of five boron atoms linked to nine hydrogen atoms. Decaborane is a toxic white crystalline solid at room temperature, melting and subliming at 100°C. Decaborane consists of 10 boron atoms linked to 14 hydrogen atoms. It can be handled in air without igniting, although it did oxidize in air at a rate dependant on the temperature, becoming reliably pyrophoric at 100°C.

Pentaborane was HEF-1 (high-energy fuel-one). Propyl pentaborane was HEF-2, ethyl decaborane was HEF-3, and methyl decaborane was HEF-4. HEF-3 and HEF-4 were reliably nonpyrophoric at room temperature. Each of the alkylated boranes was primarily monoalkylated, but each also contained some di-, tri-, and tetra-alkylated boranes. The fuels made by direct alkylation of diborane had somewhat different compositions, but were required to meet the same specifications as the materials made by alkylating pentaborane or decaborane.

In a period of four years, 1956 through 1959, five commercial plants were built to produce alkylated boranes:

Niagara Falls, New York—$5.5 million cost to produce 100 tons per year HEF-2
Lawrence, Kansas—$4 million cost to produce fuel equivalent to HEF-2
Model City, New York (Navy)—$4.5 million cost to produce 240 tons per year HEF-2
Model City, New York (Air Force)—$45 million cost to produce 5 tons per day HEF-3
Muskogee, Oklahoma (Navy)—$38 million cost to produce 5 tons per day HEF-3 equivalent

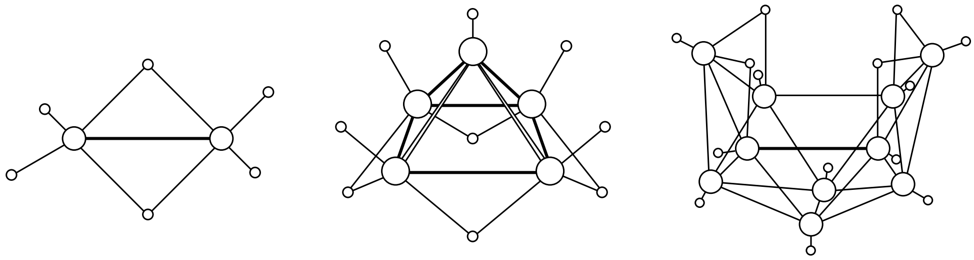
The first three plants were completed and went into operation before June 1958. Including the pilot plants, $100 million dollars were spent on plant construction, and even the largest plants listed above were considered small plants. If approved for use in the supersonic B-70 bomber, each plane could burn 20 tons per hour of HEF-3 without the afterburners and 80 tons per hour with the afterburners.

The toxic and pyrophoric nature of the boron hydrides made them difficult and dangerous to work with. All plant personnel were required to carry a gas mask and to put it on before performing any work with the boron hydrides. Many people did experience boron hydride poisoning, even though the strong odor of the boranes did provide ample warning of exposures. Fires were frequent and some explosions occurred. A total of eight people were killed in five of the accidents. The perseverance of the people on the project under these circumstances amply demonstrates their dedication to the success of the project.

During the period from 1955 to 1959, news of the project swelled from a trickle to a flood in technical publications and newspapers and became more and more accurate. Many contractors were vying for "a piece of the action" in what seemed to be the infancy of a new fuel industry. The launch of the Russian Sputniks in the fall of 1957, followed by an authoritative but false claim that the Russians had used boron fuels, helped to bring the enthusiasm to its high point in 1958. Although Russia was working on boron fuels, both the U.S. and Russian programs were driven in part by intelligence reports of the other side's efforts.

Why then, did the U.S. government cancel the project in August 1959, just when the two largest plants under construction were nearing completion? Basically, too much reliance had been placed on heat of combustion as a measure of how well the boron fuels would perform. The crucial fact was that boric oxide was produced in the combustion. In jet engines, the boric oxide fouled the rotating fan blades. In rocket engines, a solid combustion product cannot contribute to thrust as gaseous combustion products do. Hence, the boron fuels were unsuitable for the intended uses.

Curiously, the people doing rocket fuel tests probably knew the truth as early as 1957. This was evident in impractical suggestions that the boron fuels be burned with fluorine rather than oxygen, or be burned in combination with hydrogen. It seems that nobody wanted to be the first to blow the whistle. The project also had at least two fringe benefits: it was an excellent training ground for many chemists and engineers, and the chemistry of boron received a major boost in interest and understanding. Nobel prizes were awarded in 1976 to William Lipscomb for his work on the structure of boranes and in 1979 to Herbert C. Brown for his development of boron reagents important to organic synthesis. Bonding and structure of the boranes is quite unusual (see Figure 3). In the early 1940s, Herbert Brown had demonstrated the hydrolysis of sodium borohydride as a means of generating low-pressure hydrogen. The conclusion of World War II ended the construction of a sodium borohydride plant being built for that purpose, but the same process was considered at the close of the twentieth century as a source of low-pressure hydrogen for fuel cells to power the next generation of automobiles. Additional work is being done on carborane



(a) Diborane, B_2H_6 (b) Pentaborane, B_5H_9 (c) Decaborane, $B_{10}H_{14}$

Figure 3. Structural diagrams of (A) diborane, (B) pentaborane, and (C) decaborane. The large circles represent boron atoms; the small ones represent hydrogen atoms. Note the unusual three-dimensional structures of the higher borane molecules and also the unusual "bridge" hydrogen atoms, which are shared by two boron atoms.

polymers, though none are yet offered commercially. The technical literature on boron chemistry has swelled immensely.

See also **Rocket Propulsion, Liquid Propellant**

ANDREW DEQUASIE

Further Reading

Dequasie, A.E. *The Green Flame*. American Chemical Society, Washington D.C., 1991.

Holtzman, R.T. *Production of the Boranes and Related Research*. Academic Press, New York, 1967.

Stock, A. *Hydrides of Boron and Silicon*. Cornell University Press, Ithaca, 1933.

Breeding, Animal: Genetic Methods

Modern animal breeding relies on scientific methods to control production of domesticated animals, both livestock and pets, which exhibit desired physical and behavioral traits. Genetic technology aids animal breeders to attain nutritional, medical, recreational, and fashion standards demanded by consumers for animal products including meat, milk, eggs, leather, wool, and pharmaceuticals. Animals are also genetically designed to meet labor and sporting requirements for speed and endurance, conformation and beauty ideals to win show competitions, and intelligence levels to perform obediently at tasks such as herding, hunting, and tracking.

Prior to the twentieth century, humans carefully chose and managed animals with desired qualities for breeding. Unaware of genetics, people relied for centuries on observation, experience, and chance to breed selectively livestock and pets that displayed valued traits such as sturdiness, gentle temperaments, and coat colors and textures with the expectation that their offspring might also have those characteristics. Breeding of specific lineages, which produced exceptionally vigorous animals, was recorded by breeder associations as ideal specimens, honored with awards, and promoted. Animals that lacked prized assets were removed from breeding herds. Males were castrated. Sometimes culled animals were slaughtered. In eighteenth century England, Robert Bakewell initiated many fundamental animal-breeding concepts such as keeping breeding records with accurate pedigrees, evaluating young male animals with progeny tests, emphasizing family attributes with careful inbreeding, and breeding only the best animals.

During the early twentieth century, geneticists recognized the potential of agricultural genetics. Applying scientific principles from Mendelian plant breeding experiments to animal breeding, scientists developed methodologies to manipulate some of the 30,000 to 40,000 genes in farm animals to improve such factors as growth rates, fur quality, and milk production. At Iowa State University in the U.S., animal husbandry professor Jay L. Lush (1896–1982) pioneered quantitative biometrics techniques in the 1930s and 1940s. He is recognized globally as the founder of modern animal breeding and genetics based on statistical analysis. Because of Lush, an international animal breeding study center was established at Iowa State. Such research initiated the transition of animal breeding from an amateur to primarily professional activity.

Line-breeding programs for cattle began in the U.S. in 1934. Breeding strategies are pursued because breeders aspire to develop high-quality animals, which earn higher market prices, produce greater yields, receive larger sporting purses, or win more prestigious competitions. These victories and performance tests, which assess yields and growth

rates, validate genetic breeding programs. Breeders strive to create efficiently and consistently uniform types of animals that appeal to consumers. Purebred line breeding enables breeders to develop and rely on breeding stock from lineages known to produce certain characteristics such as meatier beef cattle, faster horses, stronger oxen, or tastier swine. Sheep breeders use genetics to achieve desired fleece pigment, weight, and fiber diameter, curvature, and durability. These traits are reinforced when genetically similar animals that share ancestors are bred. Inbreeding involves mating closely related animals such as siblings in an effort to emphasize genetic traits in offspring. However, recessive genes associated with undesired characteristics that breeders cannot visually detect in the parents are sometimes paired during fertilization, resulting in inferior offspring.

In contrast, outcrossing involves mating unrelated animals representing the same breed. Breeders select animals that display specific qualities, hoping that the offspring will demonstrate all of the targeted traits. Animals of different breeds within a species are crossbred to enhance strengths associated with each breed. Sometimes breeders cross animals representing varying species such as a goat and sheep producing a geep. Most cross-species hybrids are artificially created, although some occur in the wild. Mules, the offspring of a horse–donkey cross, are sterile and considered more reliable and stronger workers than their parents.

Dog hybridization commonly occurs and historically accounts for the creation of most modern breeds such as Yorkshire terriers, which were derived from an amalgamation of breeds. Mongrels are the most familiar canine hybrids. Some breeders intentionally cross breeds in an attempt to emphasize and improve breed qualities in offspring. Hybrids usually inherit their parents' best physical, or phenotype, characteristics and lack any genetic material that might cause deficiencies. They are considered more vigorous and resistant to genetic defects and diseases than purebred dogs and tend to be nonshedding, thus nonallergenic. Cockapoos, a cross between Cocker Spaniels and Poodles, are popular because of their gentle dispositions and attractive appearances. Some people purposefully mate dogs with wolves or coyotes in an attempt to create aggressive hybrids for protection. Some natural genetic mutations occur such as the Munchkin cat breed, which has short legs as a result of a dwarfism gene.

Breeding agendas changed during the twentieth century due to industrialized agriculture and cultural attitudes regarding nutrition. Through World War II, fatty swine were valued for the lard they produced which had many practical applications. Breeders cultivated hogs genetically prone to produce large amounts of fat until the postwar urban market shifted to prefer lean pork. From the 1950s, fatty swine were culled from breeding programs, and many pig breeds became extinct. As breeding became more selective to meet consumer demands, animals became physically and genetically more uniform. Such uniformity was also essential for animals to be processed by automated slaughter machinery. Animals' genetic diversity essential for sustainable agriculture was threatened as more breeds became extinct. Groups such as the American Livestock Breeds Conservancy formed with the aim of preserving animal breeds. Members are encouraged to register their animals and store sperm in semen banks. Heirloom animals, representing endangered breeds from past centuries, are protected on farms where breeding efforts strived to replenish stock. Some consumers protest at the use of genetics to control commercial animal reproduction and question how animal welfare and product quality are affected.

By the late twentieth century, genetics and mathematical models were appropriated to identify the potential of immature animals. DNA markers indicate how young animals will mature, saving breeders money by not investing in animals lacking genetic promise. Scientists also successfully transplanted sperm-producing stem cells with the goal of restoring fertility to barren breeding animals. At the National Animal Disease Center in Ames, Iowa, researchers created a gene-based test, which uses a cloned gene of the organism that causes Johne's disease in cattle in order to detect that disease to avert epidemics. Researchers also began mapping the dog genome and developing molecular techniques to evaluate canine chromosomes in the Quantitative Trait Loci (QTL). Bioinformatics incorporates computers to analyze genetic material. Some tests were developed to diagnose many of several hundred genetic canine diseases including hip dysplasia and progressive retinal atrophy (PRA). A few breed organizations modified standards to discourage breeding of genetically flawed animals and promote heterozygosity.

See also **Artificial Insemination and *In Vitro* Fertilization (Farming); Farming, Growth Promotion; Genetic Engineering, Applications**

ELIZABETH D. SCHAFER

Further Reading

Bixby, D.E., Christman, C.J., Ehrman, C.J. and Sponenberg, D.P. *Taking Stock: The North American Livestock Census*. American Livestock Breeds Conservancy, Pittsboro, 1994.

Bourdon, R.M. *Understanding Animal Breeding*, 2nd edn. Prentice Hall, Upper Saddle River, 2000.

Cameron, N.D. *Selection Indices and Prediction of Genetic Merit in Animal Breeding*. CAB International, Oxford, 1997.

Clark, A.J., Ed. *Animal Breeding: Technology for the 21st Century*. Harwood, Amsterdam, 1998.

Comstock, R.E. *Quantitative Genetics with Special Reference to Plant and Animal Breeding*. Iowa State University Press, Ames, 1996.

Maijala, K., Ed. *Genetic Resources of Pig, Sheep, and Goat*. Elsevier, New York, 1991.

Padgett, G.A. *Control of Canine Genetic Diseases*. Howell Book House, New York, 1998.

Scott, J.P. and J.L. Fuller. *Genetics and the Social Behavior of the Dog*, reprint edition. University of Chicago Press, Chicago, 1998.

Willis, M.B. *Practical Genetics for Dog Breeders*. Howell Book House, New York, 1992.

Breeding, Plant: Genetic Methods

The cultivation of plants is the world's oldest biotechnology. We have continually tried to produce improved varieties while increasing yield, features to aid cultivation and harvesting, disease, and pest resistance, or crop qualities such as longer postharvest storage life and improved taste or nutritional value.

Early changes resulted from random cross-pollination, rudimentary grafting, or spontaneous genetic change. For centuries, man kept the seed from the plants with improved characteristics to plant the following season's crop. The pioneering work of Gregor Mendel and his development of the basic laws of heredity showed for other first time that some of the processes of heredity could be altered by experimental means.

The genetic analysis of bacterial (prokaryote) genes and techniques for analysis of the higher (eukaryotic) organisms such as plants developed in parallel streams, but the rediscovery of Mendel's work in 1900 fueled a burst of activity on understanding the role of genes in inheritance. The knowledge that genes are linked along the chromosome thereby allowed mapping of genes (transduction analysis, conjugation analysis, and transformation analysis). The power of genetics to produce a desirable plant was established, and it was appreciated that controlled breeding (test crosses and back crosses) and careful analysis of the progeny could distinguish traits that were dominant or recessive, and establish pure breeding lines. Traditional horticultural techniques of artificial self-pollination and cross-pollination were also used to produce hybrids. In the 1930s the Russian Nikolai Vavilov recognized the value of genetic diversity in domesticated crop plants and their wild relatives to crop improvement, and collected seeds from the wild to study total genetic diversity and use these in breeding programs. The impact of scientific crop breeding was established by the "Green revolution" of the 1960s, when new wheat varieties with higher yields were developed by careful crop breeding. "Mutation breeding"—inducing mutations by exposing seeds to x-rays or chemicals such as sodium azide, accelerated after World War II. It was also discovered that plant cells and tissues grown in tissue culture would mutate rapidly. In the 1970s, haploid breeding,

Figure 4. To increase the genetic diversity of U.S. corn, the Germplasm Enhancement for Maize (GEM) project seeks to combine exotic germplasm, such as this unusually colored and shaped maize from Latin America, with domestic corn lines. [*Photo by Keith Weller. ARS/USDA*].

which involves producing plants from two identical sets of chromosomes, was extensively used to create new cultivars. In the twenty-first century, haploid breeding could speed up plant breeding by shortening the breeding cycle.

Although the ability to make genes change by inducing mutation and selecting beneficial varieties was well known, the development of a crop or plant with desired characteristics was lengthy, haphazard, and difficult when done by these conventional breeding techniques.

In the 1970s, recombinant DNA technology made it possible to manipulate sequences of DNA and to combine DNA from two or more sources. To be useful, the recombinant molecule must be replicated many times. Gene cloning provided a new dimension to plant biotechnology by enabling intentional directed changes to the genotype of a plant using enzymes to identify, remove, invert, and splice genes and by using mobile pieces of DNA (plasmids) to carry genes from cell to cell. Along with the knowledge of genes, understanding of the biochemistry of cells has been critical. For example knowing how genes make use of the information stored in their DNA to produce proteins and how they switch genes on and off allows us to manipulate gene expression.

One of the most important food crop plants is wheat, a plant that is difficult to modify genetically since it has a huge genome (17,000,000 kilobases) or six times more than a human cell. Because the rice genome (40,000 kilobases) is structurally similar to parts of wheat, and rice is also an important staple crop of inherent commercial importance, a novel strategy has been developed. The similar genes are identified in rice, and this can be used as a model to locate the genes for wheat.

By 2000, researchers had established the complete gene sequence (genome) for the first plant (*Arabidopsis thaliana*, a small plant in the mustard family) and are beginning to understand how genes switch on and off. This information will allow studies of how cells respond to external stimuli. To do this, it is necessary to study the total protein content of a cell (the proteome) as it changes. Genetic engineers can now access gene-sequencing data readily. It is not published in conventional scientific journals but in one of three databases that includes information from around the world. These databases can be searched for the information the plant breeder needs.

Plants can be modified in a directed way by gene addition (cloning) or gene subtraction (genes are removed or inactivated). Plants are now engineered for insect resistance, fungal resistance, viral resistance, herbicide resistance, changed nutritional content, improved taste, and improved storage. A significant commercial example is the development of Bt maize. Maize crops can be devastated by the European corn borer (*Ostrinia nubiliasis*), which tunnels into the plant and thus evades insecticide sprays. The gene product that was introduced to the plant was from a bacterium producing *Bacillus thuringiensis* toxin (Bt toxin), a substance known as a δ-endotoxin. This toxin is a protein that is 80,000 times more toxic to insects than organophosphate insecticides. Its advantage as an insecticide is that the protein accumulates in the bacterium in the form of an inactive precursor and can be introduced into the plants without any effect. When the maize is eaten by the insect, its digestive enzymes attack the protein, the resulting small protein molecules bind to and damage the insect gut, and it starves to death.

Early work to produce Bt maize was carried out with the CryllA (b) version of the protein. This contains 1115 amino acids, but the toxic activity results from the segment of amino acid from 29 to 607. Rather than try to isolate the gene from the bacterium, scientists made a length of DNA by artificial gene sequencing (putting sequences of bases together in a desired order). This allowed them to modify the way the gene would be expressed in maize. The synthetic DNA was placed into a vector (a piece of DNA used to transport it), and this was then forced into maize embryos. The embryos were grown into mature plants, and the individual plants were tested to be sure the new gene was present and to determine its activity. Other plants that have been engineered to produce δ-endotoxin include rice, cotton, potatoes, and tomatoes.

Legumes (such as cow pea and beans) have been engineered to produce proteinase inhibitors. This strategy is particularly effective against larvae that feed on weeds. Other gene addition projects have led to plants that are able to withstand fungi, bacteria, and viruses or to resist the toxic effects of herbicides used to control weeds.

Gene deletion or inactivation is also used to genetically modify a crop plant. The first genetically modified (GM) food to be approved for sale in the UK was a tomato. This was developed with antisense technology in which a gene is removed from the chromosome and reinserted in the reverse direction so that it does not function. The gene in question was the polygalacturonidase gene. Inactivation or partial inactivation of this gene delays ripening, enabling the grower to leave the

fruits on the plant until they ripen and develop a full flavor while preventing softening that makes transportation difficult. Other characteristics that result from gene subtraction include the prevention of discoloration by removal or reduced expression of the polyplenol oxidase or a lowered starch content by reduction of starch synthetase. The first GM plants were engineered for such traits, but breeders are increasingly developing engineered traits that improve nutritional content by increasing the proportion of essential amino acids, reducing the starch content, increasing the vitamin content, or adding higher monosaturated, or unsaturated, fatty acid content. One of the best examples of this technology is the development of golden rice, which contains the provitamin A gene. Rice is a staple diet in many parts of the world where a major problem is childhood blindness caused by vitamin A deficiency. Consumption of golden rice will prevent this serious condition. Because the beta-carotene, or provitamin A, is expressed as a golden color, identification of the GM variety of rice is possible.

Genetic engineering of our food supply has generated much debate, strong views, and heated argument. Our ability to change the genetic makeup of cells is seen by some as interfering with life itself. There is also concern about the possible impact of GM crops on the food supply chain and farming practices around the world, while there are powerful counterarguments that without GM crops wide-scale starvation will be a fact of life for many people. Many have questioned the motives of the multinational companies that have developed herbicide resistant crops while also manufacturing the herbicide. From a scientific point of view the important issue is whether GM crops are safe. This question can be answered both from a food safety and an environmental safety perspective. In the context of food safety, all GM crops to be marketed in Europe must pass a stringent assessment as defined by the European Novel Food Regulation (258/97). The food is assessed rigorously by leading experts who look at and evaluate complex scientific evidence. Details of the genetics, intended use, nutritional aspects, toxicology, and potential allergenicity are all scrutinized. It is true to say that we know far more about the GM food crops that have been through this process than we do about the so-called conventional agricultural crops that have resulted from years of hybridization and selection. We do however have a tradition of consumption and recognize that any potential problems might cause a serious allergic response in some sensitive people, as has been the case with peanuts. Today there is a great deal of information available relating to the development and approval of GM crops, but public reaction, particularly in Europe, has led to the withdrawal of foods containing GM ingredients from some supermarkets. Certainly the use of antibiotic-resistant markers, which were once used to identify the transformed plants during the cloning process, is no longer tolerated since there is a very small theoretical risk that the antibiotic resistance could, upon consumption of the food, be transferred to the cells in the flora of the human gut. However, it is true to say that despite consumption in the U.S. of significant amounts of GM foods, no evidence of harm has ever been reported.

Another area of concern is potential harm to the environment from GM foods. Predicting the effects that GM plants might have on the soil or the flora and fauna in the environment can be done on model systems, but ultimately there is a need for field scale evaluations to measure potential environmental effects of GM foods as compared to the non-GM crop over several seasons in a variety of locations. Until such experiments are concluded, there is a moratorium on the commercial growing of GM crops in the U.K. It is important that these field scale evaluations (which are only done after carefully controlled releases) go ahead and that the results are evaluated robustly and peer reviewed. Evidence of safe consumption and lack of environmental hazard along with public education and total transparency are required if the public is to accept GM products in agriculture.

See also **Biotechnology; Food, Processed and Fast; Genetic Engineering, Applications; Pesticides**

JANET M. BAINBRIDGE

Further Reading

Feitelson, J.S, Payne, J. and Kim ?. *Bacillus thuringiensis* and beyond. *Biotechnology*, 10, 271–275, 1992.
Smith C.J.S. *et al. Nature* 334, 724–726, 1994.
Advisory Committee on Novel Foods and Processes, Annual Report, 2000, FSA/0013/0301.

Bridges, Concrete

A complex interplay between societal change, the development of the internal combustion engine, and the impact of World War I, led to an explosion in the number of road vehicles in the immediate postwar years—and a totally inadequate nineteenth century legacy of roads to accommodate them. Following the first International Road

Congress in 1923, vast and expensive road-building programs were undertaken in the U.S. and Europe, particularly in Germany, during the 1930s. After World War II highway construction continued to grow in an attempt to keep pace with the popularity of the car for private transport. Concrete—strong in compression but weak in tension—is not particularly satisfactory as a running surface. It can easily crack, unlike tarmac, though in the 1960s its use as a surface did become widespread. Otherwise, however, concrete became omnipresent in twentieth century road construction, and in the myriad of bridges, large and small, associated with highway networks.

Because of unreinforced concrete's limitations, nineteenth century concrete bridges were built as arches, following the usual form of bridges in natural stone, which has similar strength and weakness. Paradoxically, it was the incorporation of the "rival" and far more expensive material steel—equally strong in compression but also with high tensile strength—that enabled concrete thus reinforced to become the twentieth century's most widely used material for bridges carrying road and sometimes rail traffic.

The French engineer François Hennebique notably developed techniques for embedding steel reinforcing rods in concrete for structural effectiveness, and had built around 100 reinforced concrete bridges by 1900. Already his designs were exploiting the material's potential for strength and slenderness in arch designs that were much thinner than was possible with stone or mass concrete; but the first, and arguably still the greatest, designer of reinforced concrete bridges was the Swiss, Robert Maillart. Though Maillart only built in his native country and despite the fact that none of his spans exceeded 90 meters, the elegance, economy, and inventiveness of his designs, combined with their fitness for purpose and for their often sublimely beautiful locations, made him one of the most influential twentieth century bridge designers following his death in 1940.

Even more than simple reinforcement, prestressing opened up new possibilities in concrete bridge building. As so often with technical innovation, the concept—of tightening metal rods or strands within concrete to further increase its tensile strength—had been around for a long time before it became a practical proposition. That it did so is due to the observation and tenacity of another French engineer, Eugene Freyssinet, who around 1910 first observed the tendency of concrete to "creep" (its continuing slow shrinkage after solidification) on his first reinforced concrete bridges. In later projects he introduced the practice of jacking the arch halves apart after casting, and inserting extra concrete at the crown between them to counter the effects of the shrinkage. The scale in particular of his Plougastel Bridge, completed in Brittany in 1930 and then by far the world's largest reinforced concrete bridge, made it necessary for him to study and evaluate the effects of creep as exactly as possible. His researches led him to the view that "locking in" tensile strength by incorporating steel strands in the concrete, stretched to a precisely calculated extent, was a viable structural system—and indeed would effectively produce a new building material.

Freyssinet's first prewar attempt to mass-produce and market prestressing failed, but after World War II he successfully built six single-span prestressed concrete bridges across the River Marne in France. Prestressed concrete, either pre- or posttensioned, rapidly became the material of choice for some short-, most medium-, and more rarely, some long-span bridges. In pretensioning, the concrete is poured around tendons that have already been stressed against an anchor frame, this being released when the concrete has hardened so that the tensile strength locked into the strands is imparted to the concrete adhering to them. In posttensioning, the strands are threaded through voids cast into already-hardened concrete, and then tightened.

Reinforced concrete and prestressed concrete are used in several structural forms in modern bridge building. For most short single spans, the concrete is cast as an arch or a solid slab *in situ* (literally "on site") on formwork and around the mesh of reinforcing rods. For simple spans from around 16 to 20 meters, the reinforced slab is usually cast with voids to lighten the weight of the concrete. Prestressing is normally introduced when longer spans are required. Such bridges can be anything from a single span across a road to the literally thousands that comprise the 38-plus kilometer Lake Pontchartrain Bridges in Louisiana. Beams, of rectangular or T-shaped cross-section, are prestressed and precast offsite, craned into place on supporting piers, and topped with deck slabs. For yet longer spans, sections of prestressed concrete box girder (see the entry "Bridges, Steel") may be are joined together between supports of up to 200 meters and more. These bridges often have the appearance of a wide, shallow arch, though rarely do they act structurally as a true arch, in which forces are carried around and down the arch and into abutments. Instead, the structure acts as a beam, with gravity creating compression forces

along the top of the span and tension forces along the bottom, which the prestressing withstands.

The longest-span wholly prestressed concrete bridges are, however, true arches. The archetype is the 305-meter-span Gladesville Bridge in Sydney, completed across the Parramatta River in 1964. Its design was unusual, in that it followed the same voussoir principle that the Romans used for their masonry arches, in which wedge-shaped stones were cut to form segments of a semicircle. In the case of the Gladesville Bridge, the voussoir units are hollow prestressed concrete boxes, each precast in the shape necessary to form the giant shallow arch of the main bridge structure, from the upper surface of which the even shallower curve of the precast road deck is carried on slender upright prestressed piers.

Even longer concrete spans do exist in which a concrete deck or concrete pylons may form part of a cable-stayed or a suspension bridge. However, as all bridges in these forms also incorporate steel, always in the hangers of both types and in the cables of suspension bridges, they are discussed in a separate entry, as are all-steel suspension bridges built in the early twentieth century.

See also **Bridges, Long Span and Suspension; Bridges, Steel; Concrete**

DAVID J. BROWN

Further Reading

Brown, D.J. *Bridges: Three Thousand Years of Defying Nature*. Mitchell Beazley, London, 1993. A well-illustrated popular history.

Concrete Bridge Development Group. *Durable Bonded Post-Tensioned Concrete Bridges*. Concrete Society, London, 1996.

Dekenkolb, O.H. and American Concrete Institute. *Concrete Box Girder Bridges*. State University Press, Iowa, 1977.

Gies, J. *Bridges and Men*. Cassell & Company, London, 1964. No longer up-to-date, but an excellent popular narrative of bridges and their designers.

Heins, C.P. and Lawrie, R.A. *Design of Modern Concrete Highway Bridges*. Wiley, Canada, 1984.

Leonhardt, F. *Bridges: Aesthetics and Design*. Architectural Press, London, 1982. A seminal work on bridge design by one of the greatest twentieth century bridge engineers.

Ryall, M., Peake, G. and Harding, J., Eds. *Manual of Bridge Engineering*. Thomas Telford, London, 2000. A definitive text on the design and construction of all types of modern bridge, published under the auspices of the Institution of Civil Engineers.

Wallbank, E.J., Department of Transport and G. Maunsell and Partners. *The Performance of Concrete in Bridges: A Survey of 200 Highway Bridges*. Her Majesty's Stationery Office, London, 1989.

Useful Websites

http://www.bridgeweb.com/

Bridges, Long Span and Suspension

From the beginning of the twentieth century, bridge spans in excess of 300 meters became increasingly common. Depending on considerations of location, use, and loading—not to mention aesthetic and engineering aspiration—these could be suspension, arch, or cantilever structures. When spans of 1000 meters or more began to be contemplated from around 1930 however, a suspension bridge was the only answer. The breakthrough structure was New York's George Washington Bridge; its clear span of 1067 meters almost doubled that of the previous record-holder, the 564-meter Ambassador Bridge in Detroit completed only two years earlier. Nonetheless, within a few years the leading edge of enterprise had passed to the West Coast, with the simultaneous construction of the San Francisco Bay Bridge complex (twin 704-meter suspension spans plus a tunnel and a cantilever), and the 1280-meter-span Golden Gate Bridge, opened in 1937.

The Bay Bridge was designed under the supervision of California State highway supremo, Charles H. Purcell, while the Golden Gate was the one great work of the engineer Joseph Strauss. Others however, like Othmar Ammann, Leon Moisseiff, and David Steinman, were signature designers who left a monumental legacy in steel across the length and breadth of the U.S.

In the design of the many large suspension bridges constructed in the burst of activity in the 1930s, stability of the unprecedented extended decks was a primary concern. The response of long, narrow unsupported structures to wind forces was not precisely understood, though the steel plate girder deck of the George Washington Bridge was correctly deemed by its designer, Ammann, to be wide (36.3 meters) and thick (3 meters) enough to be stiff and stable without a supporting truss. The longer and narrower Golden Gate, by contrast, was given a 7.6-meter-deep stiffening truss, whilst the need for the Bay Bridge to carry two levels of traffic necessitated a 9-meter-deep truss to carry the second deck. Other long-span bridges for which far lower volumes of traffic were anticipated, however, invited a combination of narrowness and slenderness in the pursuit of elegance which proved ultimately disastrous when Moisseiff's Tacoma Narrows Bridge ("Galloping Gertie") collapsed in a wind of only 68 kilometers per hour

(km/h) in November 1940. Like that of the George Washington Bridge, its deck was a straight-sided plate girder, but only 11.9 meters wide and 2.4 meters thick over its 853-meter length, and its designed-in flexibility led to twisting under quite moderate winds that progressively increased under its own momentum until failure occurred.

For the next 25 years no long-span suspension bridge was designed without a stiffening truss, including the magnificent 1158-meter-span Mackinac Straits Bridge between Great Lakes Huron and Michigan—the crowning masterpiece of David Steinman's career—and the 1298-meter-span Verrazano Narrows Bridge in New York. This, the last great work of Othmar Ammann, took the world record span from the Golden Gate in 1964 and held it for 17 years.

The challenge of designing a slender, untrussed but aerodynamically stable deck was finally met successfully by a team of British engineers led by Ralph Freeman when the 987.5-m-span Severn Bridge in England was given an "aerofoil" with trailing edges tapered in cross-section on both sides. This minimized wind resistance and avoided the buffeting eddies of wind created by the square sides of the earlier American plate girder designs. Aerodynamic decks became the norm for a new generation of large suspension bridges in Europe, including the two Bosporus Bridges in Turkey (1074 meters and 1090 meters), the record-breaking 1410 meter span Humber Bridge in England, opened in 1981, and the even larger East Bridge across the Great Belt in Denmark, completed in 1998 with a main span of 1624 meters.

East Asia, however, saw the late-twentieth century's greatest activity in long-span bridge design and construction, the most intensive building program taking place in Japan with three major links comprising over 20 individual bridges completed between the islands of Honshu and Shikoku in the 1980s and 1990s. These included no fewer than eight large suspension bridges, among them the world-record holder at the end of the century, the Akashi Kaikyo, opened in 1998 with a main span of 1991 meters.

Suspension bridges have their decks literally suspended by hangers (usually vertical) from thick steel cables formed from thousands of steel strands spun back and forth between the tops of the pylons (towers) and then compacted and wrapped. This technique was pioneered in the nineteenth century by John Roebling, most famous as the designer of the Brooklyn Bridge in New York. Cable-stayed bridges, equally literally, have their decks "stayed" by being directly connected to supporting masts with straight cables. Beyond this, there are many variables in design: pylons single or double or configured as a A-shape or an inverted Y; and anything from a single cable connecting pylon and deck to many, either fanning out from near the top of the pylon or parallel to each other. The first modern cable-stayed bridges were three in number, designed and built in close proximity to each other in 1952 in Düsseldorf, by the great German engineer Fritz Leonhardt.

In the remainder of the century economy and esthetics made the cable-stayed form the design of choice for the majority of spans upward of 300 meters in most parts of the world. Two landmark structures, completed respectively in 1994 and 1998, showed cable staying being used on a scale for which only suspension spans would previously have been conceived. These were the 856-meter-span Pont de Normandie in northern France and the 890-meter-span Tatara Bridge, yet another element in Japan's great Honshu–Shikoku complex. As the century closed, cable-stayed spans over 1 kilometer were being actively planned, though the suspension concept envisaged for the long-dreamt-of bridge between Italy and Sicily, at 3.3 kilometers, still far exceeds any possible extension of the cable-staying principle.

See also **Bridges, Concrete; Bridges, Steel**

DAVID J. BROWN

Further Reading

Brown, D.J. *Bridges: Three Thousand Years of Defying Nature*. Mitchell Beazley, London, 1993. A well-illustrated popular history.

Gies, J. *Bridges and Men*. Cassell & Company, London, 1964. No longer up-to-date, but an excellent popular narrative of bridges and their designers.

Gimsing, N.J. *Cable Supported Bridges: Concept and Design*, 2nd edn. Wiley, Chichester, 1997. A definitive study by one of the late twentieth century's most distinguished designers.

Leonhardt, F. *Bridges: Aesthetics and Design*. Architectural Press, London, 1982. A seminal work on bridge design, by one of the greatest twentieth century bridge engineers.

Pugsley, A. *The Theory of Suspension Bridges*, 2nd edn. Edward Arnold, London, 1968.

Ryall, M., Peake, G. and Harding, J., Eds. *Manual of Bridge Engineering*. Thomas Telford, London, 2000. A definitive text on the design and construction of all types of modern bridge, published under the auspices of the Institution of Civil Engineers.

Useful website

http://www.bridgeweb.com/

Bridges, Steel

Though techniques for smelting steel had been known in principle since antiquity, only from the mid-nineteenth century did its large-scale production as a practical structural material become a reality. Stronger than wrought iron and more ductile than cast iron, its superior qualities were exploited in three great steel bridges, each in a different structural system, built between 1870 and 1890. The triple-arch St. Louis Bridge in Missouri, with its two levels for road and rail, the suspension Brooklyn Bridge in New York, and the double-cantilever Forth Rail Bridge in Scotland neatly prefigured the resourcefulness with which twentieth century bridge engineers would continue to exploit the material in long-span structures. With growing understanding of the structural potential of steel, and improvements in its tensile strength and other properties, bridges continued to progressively increase in span.

By the end of the nineteenth century a succession of major steel arch bridges had been constructed, notably in southern France, but one of the twentieth century's first such milestone structures was completed in 1907 across the Zambezi Gorge in Zimbabwe by the British engineer, Ralph Freeman. Its span of 152 meters was virtually doubled less than ten years later, when the American engineer Gustav Lindenthal completed his huge 298-meter arch truss Hell Gate Bridge across an arm of the East River, New York. This bridge held its world record for an arch span for 15 years, until the Bayonne Bridge, designed by one of Lindenthal's assistants named Othmar Ammann, was completed across the Kill Van Kull, also in New York, in 1931. Unlike the rail Hell Gate Bridge, the Bayonne was a much lighter structure for road traffic only, and at span of 503.6 meters had been deliberately designed to just exceed a mighty rival, the Sydney Harbor Bridge (Figure 6). Begun four years earlier but completed a few months later in March 1932, the latter remains the widest of all long-span bridges, built to carry both rail and roadway side by side. As steel arch structures, these two giants on opposite sides of the world were only exceeded in 1978 when the 518-meter-span New River Gorge Bridge was completed in West Virginia. Unlike its three predecessors, whose decks are suspended from the arches, the New River Gorge Bridge—as its name indicates—spans a deep river valley, and therefore carries its deck atop the steel arch.

No steel arch bridge was ever the world's longest span *per se*. Cantilever bridges, on the other hand, in their time were world-beaters. In these, arms extend from both sides of rigid bases and either

Figure 5. Sfalassa Bridge.

Figure 6. Sydney Harbor.

meet in the middle or support a central span. In the case of the Forth Bridge, there had been three supporting towers and two clear 521-meter spans, double world-beaters when it was completed in 1889. Superficially similar in profile, but with a single and even longer main span of 549 meters, the Quebec Bridge in Canada took over the record when it was completed in 1917 after 13 years' work and two major construction disasters involving 85 fatalities. Steel cantilever bridges continued to be a popular structural type for large bridges, particularly in the first half of the twentieth century. Two notable examples were the Queensboro Bridge in New York (the first to be constructed in high-strength nickel steel), and the Carquinez Straits Bridges in San Francisco. For the longest spans of all, however, favor shifted to the suspension bridge with a series of constructions in the U.S. whose single greatest "leap forward" was the mighty 1067-meter-span George Washington Bridge in New York, completed in 1931. Though throughout the remainder of the century even longer-span suspension bridges continued to be designed and built (discussed in the entry Bridges, Long-Span/Suspension), the George Washington, with its truss-framed double deck and unique open 183-meter skeletal steel towers, remains perhaps the archetypal all-steel suspension bridge.

The second half of the twentieth century saw two distinct developments in bridge design: the first, cable staying, is discussed in the entry on long span and suspension bridges. The second is the box-girder. It also had nineteenth century progenitors in the iron tubular rail bridges erected by Robert Stephenson at Menai and Conway, but the post-World War II steel box girder was a far smaller and lighter affair, its design actually built on new understanding of the behavior of thin-walled boxes and tubes in torsion (twisting) derived from the design of wartime aircraft fuselages. A box girder is essentially a hollow beam formed from a series of open-ended boxes made of steel welded together, with stiffened horizontal plates or flanges at the top and bottom connected by vertical or angled side-plates or webs. Box-girder bridges were quickly found to be strong, economical in material, and relatively quick to construct, but progress in their use came to a temporary halt in 1970 when two major failures occurred within three months, one at Milford Haven in Wales and the other on the West Gate Bridge over the River Yarra at Melbourne. In both cases joints between boxes failed, at Milford Haven above a column and at the West Gate at the midpoint of a completed span. Though the collapses seemed similar, and both involved fatalities, extensive

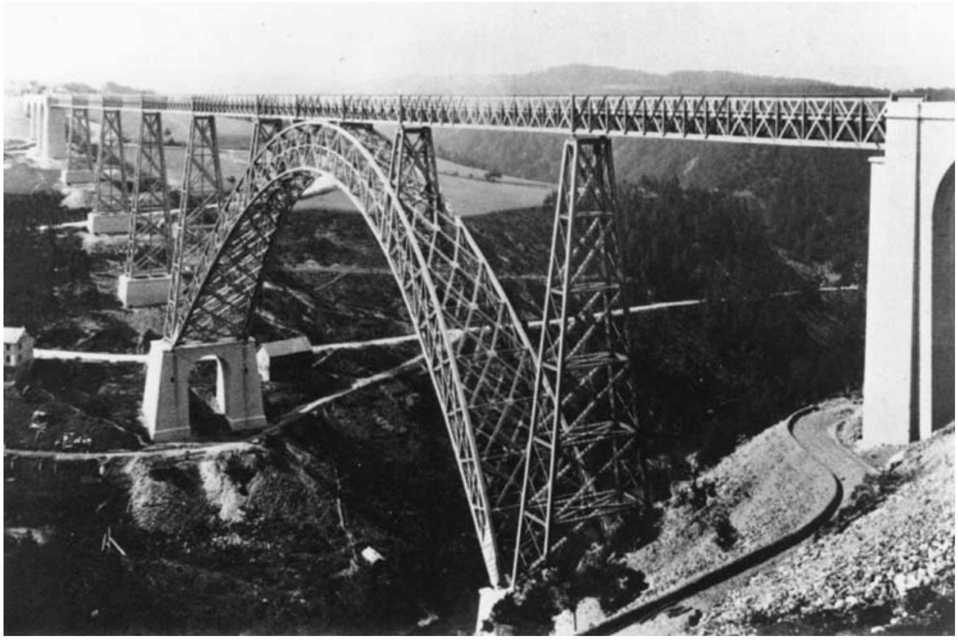


Figure 7. Garabit Viaduct.

investigations showed rather different causes: failure of a pier support diaphragm in the former and inadequacies in erection methods and site organization in the latter. Greater understanding of the forces acting within and upon the steel components led to new and much more stringent construction standards, and in the decades since, steel box girders have continued to be a widely used form of bridge construction all over the world.

Figure 8. Victoria Falls.

Steel is also very commonly used for constructing footbridges, and particularly towards the end of the century, there was a proliferation of creative invention in their design. A pair of braced girders or a small steel box remains the simplest forms of structure, whilst for a longer span a truss form will deliver a lighter and economical bridge. If a landmark structure is needed, however, the variations and combinations of—in particular—cable-stayed and arched designs seems virtually limitless. A notable inspiration for other designers has been the numerous memorable bridge profiles created by the Spanish architect and engineer Santiago Calatrava. Pedestrian decks may be braced by virtually invisible stainless steel wires extending from vertical or angled masts, suspended from slender tubular arches, or as in the truly innovative Gateshead Millennium Bridge designed by Wilkinson Eyre, curved in plan and pivoted to raise in conjunction with its balancing steel arch to allow shipping to pass beneath, an action which recalls the form and action of a gigantic "blinking eye."

See also **Bridges, Concrete; Bridges, Long Span and Suspension**

DAVID J. BROWN

Further Reading

Brown, D.J. *Bridges: Three Thousand Years of Defying Nature*. Mitchell Beazley, London, 1993. A well-illustrated popular history.

Chatterjee, S. *The Design of Modern Steel Bridges*. BSP Professional Books, Oxford, 1991. A key textbook, particularly on plate and box-girder bridges.
Corus Construction Centre. *The Design of Steel Footbridges*. Corus U.K., Scunthorpe, 2000.
European Convention for Constructional Steelwork and British Constructional Steelwork Association. *Proceedings of the International Symposium on Steel Bridges, Institution of Civil Engineers, 25–26 February 1988*. ECCS, Brussels, 1988.
Gies, J. *Bridges and Men*. Cassell & Company, London, 1964. No longer up-to-date, but an excellent popular narrative of bridges and their designers.
Leonhardt, F. *Bridges: Aesthetics and Design*. Architectural Press, London, 1982. A seminal work on bridge design by one of the greatest twentieth century bridge engineers.
Matildi, P. and Matildi, G. *Ponti Metallici: Esperienze Vissute*. Costruzioni Cimolai Armando SPA, Pordenone, Italy, 1990. A useful Italian perspective on the subject, with many construction details and captions in English as well as Italian.
Ryall, M., Peake, G. and Harding, J., Eds. *Manual of Bridge Engineering*. Thomas Telford, London, 2000. A definitive text on the design and construction of all types of modern bridge, published under the auspices of the Institution of Civil Engineers.
Steel Construction Institute. *Integral Steel Bridges: Design Guidance*. SCI, Ascot, 1997.
Tordoff, D. and British Constructional Steelwork Association. *Steel Bridges*. BCSA, London, 1985.

Useful websites

http://www.corusgroup.com/
http://www.bridgeweb.com/

Building Acoustics

An important element in a properly functioning building is correct building acoustics. Achieving a low level of background noise in a classroom, for example, will ensure that the teacher's voice is audible; the sounds of an orchestra will be optimal in a concert hall with proper acoustics. The systematic study of room acoustics began at the end of the nineteenth century, and consequently a scientific understanding of building acoustic design is almost entirely a twentieth century phenomenon.

The means to achieve low noise levels in buildings were developed during the twentieth century. One of the greatest differences between old and new auditoriums is the low noise levels achieved in those built from the mid-twentieth century onward. Noise from external sources can enter a room through vibration paths (structure-borne transmission) or can pass directly into the building through adjacent walls (airborne transmission). Where very low noise or vibration levels are needed in auditoriums, recording studios, and operating theaters, vibration isolation (springs and resilient materials) are used, as are physical breaks in vibration paths. Airborne noise is reduced by the use of constructions such as double partitions separated by air gaps containing absorbent materials. The failure to achieve the desired background noise levels is often due simply to poor workmanship.

Building service equipment such as boilers, etc., should be mounted on vibration isolators, and the structure and airborne paths should be considered in the design. Ventilation outlets may require silencers; low-velocity air conditioning is favored because it is quieter. Noise generated within rooms can also be a problem. For example the twentieth century saw a great increase in the use of atria, which can be very noisy, reverberant spaces due to footfall noise and the sounds of noisy items such as escalators. Perversely, a few rooms require the application of noise in order to achieve confidentiality. Masking noise from a radio is often added to hospital waiting rooms to provide privacy for consultations.

By the end of twentieth century the most critical design requirements for correct acoustics in rooms were well established. For speech, the requirement is often for intelligibility rather than fidelity. This requires the speech to be louder than the background noise and limits the room size that can be used without electronic enhancement. The direct path between the speaker and listener should not be obstructed, and the audience should be as close as possible to the speaker; placing seats on an incline is useful. Speakers should always face the audience, however, the audience can surround the speaker in small theaters. There should be hard surfaces close to the speaker and listeners to create beneficial reinforcing sound reflections. Where the stage area of a theater is very absorbent due to the scenery, reinforcing reflections from the ceiling and proscenium arch are vital. Surfaces that generate late-arriving reflections are usually treated with absorption to prevent echoes. Many of these principles were exploited in ancient amphitheaters, but it is only in the twentieth century that the scientific reasons for good or bad acoustics were understood. These requirements are easier to achieve when the speakers are located in one place, such as the stage of a theater and difficult in courts and debating chambers where there are many different speaker positions.

The room sound should not reverberate excessively, otherwise the sound of one syllable will run into the next syllable. Acoustic absorption is used

to reduce reverberance. During the twentieth century, various technologies for testing speech intelligibility were developed. In the early 1970s, T. Houtgast and J.J.M. Steeneken reported on their studies showing a direct correlation between modulation reduction factors and speech intelligibility. Their studies are the basis for the Speech Transmission Index (STI) program, an objective measure used in performance specifications. Perceptual tests may also be carried out, but these tests are rather slow to do.

Developments in electronics have influenced building acoustics. Speech reinforcement and public address systems are used for emergency evacuation and day-to-day messaging. Electronic reinforcement will sound unnatural if the speech appears to come from the loudspeakers rather than the speaker. This is solved by applying delays to the sound coming from the loudspeaker. The Haas (precedence) effect states that the sound that arrives first—the first sound heard—will usually determine the perceived location of the sound. For a room with a high ceiling such as an auditorium, a single large cluster of loudspeakers may be used. For a room with a low ceiling, a series of loudspeakers placed along the length of the room may be used. Each loudspeaker covers a different area and has different delays. To prevent feedback, sound from the loudspeakers should not be directly picked up by the microphone. Loudspeakers can be sited forward of the speaker and directional microphones used. If a space is overly reverberant, only the frequency ranges important for speech intelligibility are reproduced. Operators need to be trained to use these sound systems correctly, otherwise the speech produced may be of poor quality.

Sound reproduction rooms such as control rooms in studios tend to be small. Low-frequency resonances of the room cause coloration, but the audible effects are reduced by appropriate choice of room dimensions and techniques for the application of resonant acoustic absorption. These consist of vibrating membranes over a cavity, with resistive material such as mineral wool in the cavity. At higher frequencies, the coloration caused by early-arriving loud reflections must be minimized. Absorbers (which remove sound energy) or diffusers (which spatially and temporally disperse sound energy) can be used. Absorber technology was developed throughout the twentieth century, while diffuser designs date only to the mid-1970s. In the 1970s, revolutionary new diffuser designs used phase gratings based on mathematical sequences; designs even exploited the emerging mathematical discipline of fractals. More modern diffusers use numerical optimization algorithms to provide curved diffusers that complement contemporary architecture.

A great concert hall acts as an extension to the musical instruments, embellishing and improving the sound produced by the musicians. In 1895–97, physicist Wallace C. Sabine, in tests aimed at correcting the poor acoustics in the lecture hall at Harvard University's Fogg Art Museum, established the need to obtain the correct reverberance. Sabine has been called the father of architectural acoustics, and was the first to apply quantitative measures to the design of a concert hall (the first auditorium that was designed by Sabine was the new Boston Music Hall, opened October 15, 1900). In the last 30 years of the twentieth century, additional parameters to reverberance have been identified as important. For example, acousticians now understand how the hall shape influences sound quality. In previous centuries, trial and error had established the 'shoebox' shape (long, high, and narrow) as providing a good sound; it is now understood that this shape worked because it provided beneficial side reflections. Circular and elliptical shapes risk focusing sound-creating "hot spots." The influence of amphitheaters, theaters, and early cinema can be seen in fan-shaped halls built in the mid twentieth century, but this shape can lead to a lack of side reflections. New shapes created for halls over the last half of the twentieth century include the Vineyard Terrace, which subdivides the audience area so the dividing walls produce beneficial early reflection. In the last decades of the century, acousticians developed a better understanding of the needs of musicians. Unless musicians receive reflections from surfaces around and above the stage, they cannot hear themselves or others, and so cannot create a good tone and blend or play in time.

Multipurpose halls create problems given the different acoustic demands of different events. The general principle is to design for a primary purpose, and then adjust the acoustics for other uses. Absorption is used to deaden a concert hall ready for electronic music, and electronic means are used to make a speech theater more reverberant for music. Electronic enhancement began with the assisted resonance system in the 1960s, and there was slow and steady growth in the use of enhancement systems in theaters from that time. Sound is picked up from the stage using microphones. The signals are then delayed and played from loudspeakers to create extra reflections and so increase reverberance.

The biggest influence that electronics has had on building acoustics has been the computer. Sophisticated computer-based instrumentation has allowed accurate measurement of building acoustics. Computer-based prediction models have enabled the improved understanding and design of acoustic technologies, from building elements to the whole rooms. Much of the mathematics used by acoustic engineers was developed in the nineteenth century, but this has only been exploitable at the end of the twentieth century using computers. There was also increased interest in virtual acoustic prototypes, which would allow building acoustics to be listened to in virtual environments, allowing nonacoustic experts to more readily understand the principles of good acoustic design.

See also **Audio Recording, Electronic Methods; Loudspeakers and Earphones**

TREVOR COX

Further Reading

Barron, M. *Auditorium Acoustics and Architectural Design*. Spon Press, 1993.

Beranek, L.L. *Concert and Opera Halls: How They Sound.* Acoustical Society of America, 1996.

Houtgast, T. and Steeneken, H.J.M. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica,* 28, 66–73, 1973.

Rayleigh, J.W.S. *The Theory of Sound,* vols. 1 & 2. Dover Classics of Science & Mathematics, 1976.

Sabine, W.C. *Collected Papers on Acoustics*. Harvard University Press, Cambridge, MA, 1922. Republished by Acoustical Society of America, 1993.

Sharland, I. *Woods Practical Guide to Noise Control*. Woods Acoustics, 1990.

Smith, B. J., Peters J. and Owen S. *Acoustics and Noise Control*, Longman, 1996.

Templeton, D., Ed. *Acoustics in the Built Environment*. Architectural Press, 1997.

Thompson, E. *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900–33*. MIT Press, Cambridge, MA, 2002.

Steeneken, H.J.M. and Houtgast, T. A physical method for measuring speech transmission quality. *J. Acoust. Soc. Am*. 67, 318, 1980.

Building Equipment, *see* **Construction Equipment**

Buildings, Designs for Energy Conservation

Obtaining our energy supplies from coal, oil or gas (the fossil fuels) results in the production of a range of pollutants that have adverse effects on people, forests, wildlife, and aquatic environments. By the last quarter of the twentieth century, increased fossil fuel consumption was believed by many to have a profound effect on Earth's climate as result of the release of so-called greenhouse gases such as carbon dioxide (CO_2).

In most countries in the twentieth century, the energy consumed in buildings represented a substantial proportion of nationwide energy consumption. In higher latitude regions, the majority of this energy demand has historically been energy for homes to provide space heating, followed by energy for hot water, for powering appliances, and for lighting. In nondomestic buildings in these regions the demand has historically been dominated by electricity for lighting, appliances, and ventilation and cooling. While space and water heating can be the largest proportion of household energy consumption, electricity consumption can be as important in terms of upstream CO_2 emissions if it is generated in a fossil fuel electricity generating station. Architects, builders, and engineers have struggled to balance the demand for energy, particularly in the industrialized countries that are heavily energy dependent, with environmental and cost concerns. The oil crisis of 1973, following an embargo of oil directed primarily against the U.S. by Middle Eastern oil-producing companies, and the OPEC oil crisis of 1979 was the end of the era of cheap energy. Energy conservation emerged as a concern for both designers and consumers, particularly in countries solely dependent on imported oil. The government in Korea, for example, asked people to "think poor," reduced the number and size of electric light bulbs in government and corporate buildings, and discouraged the use of elevators, air conditioning, and street lighting. Later policies supported use and development of energy conservation technologies. In the U.S. and also in Japan, large-scale research and development funding resulted in building guidelines and technologies for energy conservation that are discussed in this entry.

Space Heating

In residential buildings in high latitudes and other colder regions it is possible to substantially reduce or eliminate the need for space heating energy by a combination of building techniques:

- Very high standards of thermal insulation of walls, roofs, and floors
- High thermal performance glazing of windows
- Airtight construction (e.g., sealing cracks between frames and windows)
- Energy-efficient ventilation (passive ventilation or efficient mechanical ventilation)

- Efficient heat recovery in ventilation systems (a heat exchanger recovers heat from the indoor air being expelled)
- Compact building form to reduce surface area to volume ratio
- Thermostatically controlled responsive space heating delivery systems
- Use highest building density possible

In technology terms, new materials are key to thermal insulation and glazing. Cellulose insulation, in the form of sawdust and paper, has long been used in walls and ceilings. Modern blown or sprayed cellulose fiber is made from fire-retardant paper, and is popular for uninsulated wall cavities. Fiberglass has been used since the 1930s as a building insulator. Plastics developed in the 1930s were used in foam form after World War II as thermal insulation (e.g., polyurethane, polystyrene or styrofoam). Ceramic fibers, commercially available from the 1970s, are used industrially as a replacement for asbestos but are not widely used in building insulation. Low-emissivity coatings (very thin, transparent layers of metals or oxides) on glass were developed in the early 1980s to reduce heat transfer to the outside. Use of a vacuum between glass panes or filling the gap between the glass panes with low-conductivity gas such as argon or krypton at atmospheric pressure reduces conductive and convective heat transfer.

Energy demand for space heating can be further reduced if passive solar measures such as the following are included:

- Orienting the main aspect of the building and the majority of glazing to face the noon-day sun (toward the south in the northern hemisphere and toward the north in the southern hemisphere)
- Designing buildings so that the most frequently used rooms are on the side nearest the midday sun
- Including materials of high thermal capacity inside the building envelope to reduce temperature swings and absorb solar heat gains
- Locating sun spaces on building aspects facing the midday sun, provided that they are not unheated or air conditioned
- Avoiding overshading the building during the winter months

Passive solar heating through architectural design has been in use since at least the fifth century BC. Archaeological evidence of house plans and the writings of scholars such as Socrates, show that ancient Greece and Rome were well aware of passive solar design. Anasazi cliff dwellings of the American Southwest were also orientated to maximize winter sun and summer shade.

Cooling

Domestic buildings in tropical regions of the world and Middle Eastern deserts have traditionally been designed with shading and air circulation patterns in mind. The rise of industrialization and commerce places more individuals in factories and office buildings and has increased the energy needs for such settings.

In many nondomestic buildings, the energy required for lighting and equipment as well as cooling may be more significant than that for space heating. In lower latitude regions and regions subject to very high solar gains, the major energy demand has historically been for providing space cooling or air conditioning, followed by hot water provision and electricity for powering appliances and lighting. It may be possible to reduce energy demand for cooling by taking account of the following:

- Orienting the main aspect of the building to face away from the midday sun (toward the south in the northern hemisphere and toward the north in the southern hemisphere)
- Minimizing the glazing that is not solar protected on the western and eastern façades
- Maximizing the ceiling heights of the rooms
- Including north-light roof lights to maximize day lighting without incurring high solar heat gains
- Designing the building to incorporate solar shading devices to keep high temperature solar gains from entering the building
- Including materials of high thermal capacity inside the building envelope
- Incorporating lighter colors and reflective surfaces on the building to reflect solar gains from the building
- Designing the building to incorporate ventilation cooling towers or solar chimneys to achieve *passive stack cooling*
- Where appropriate considering the use of evaporative cooling techniques
- Considering the use of earth tubes to precool ventilation air
- Where feasible, considering controlling the building temperature to achieve night cooling of thermally massive components of the building to keep the building cool during the day

- Where wind speeds are sufficient, considering the use of wind energy for cooling utilizing wind ventilation towers, wind scoops, or wind catchers

Wind catchers in traditional Arabic architecture used a "chimney" with one end underground and the other set over a specific height on the roof. The air trap operated according to temperature differences, with difference in air density resulting in air flow that reversed.

Other Electrical Demands

Once the larger problems of space heating or cooling have been addressed, other electrical demands should be considered. In hot water heating, the energy required for providing domestic hot water can be minimized by switching from baths to showers and using atomizing showerheads to limit water flow instead of "power-showers." Careful selection of low water consumption and low-temperature washing machines should be used in preference to more profligate models. Hot water demand is also driven by lifestyles and the conservation awareness of householders; for example, the frequency of taking a shower or bath, doing the laundry, or using dishwashers in preference to hand-washing dishes.

All buildings used by people require lighting at certain times of day, but buildings can be designed to minimize the use of artificial lighting by maximizing the use of day lighting, providing manually controlled task lighting, and providing daylight and motion-sensing controls, which can regulate artificial lighting levels and switch off unnecessary lights when no one is in the room. Building design can employ north-facing (or south-facing in the southern hemisphere) glazed façades for tasks which require good lighting, and minimize the depth of rooms from a window to maximize the level of daylight penetration. Electricity consumption can also be minimized by avoiding the use of incandescent lamps and instead using more efficient lamps such as fluorescent tubes or compact fluorescent lamps (CFLs). Better still, light emitting diode (LED) lamps were becoming available for certain purposes by the late 1990s (see Lighting Techniques).

By the 1980s and 1990s, many countries had instituted standards for energy efficiency that applied to such household appliances as refrigerators, dishwashers, and washing machines and dryers. U.S. federal law from 1978 mandated efficiency standards for major residential energy-using equipment, including heating, cooling, and water heating equipment, but standards were not finalized until legislation in 1987. Certain states, especially California, have efficiency standards for this equipment that are more stringent than the federal standards. In 1993 the Environment Protection Agency introduced the voluntary Energy Star label program to promote energy-efficient products. Consumers were urged to select energy-efficient models for conservation and cost reasons. Another option was to use portable equipment powered by batteries, such as portable computers, which are likely designed to be more energy efficient than standard electric equipment. In 1995 Energy Star was extended to cover new homes and commercial and industrial buildings.

Obtaining Building Energy Supply from Low-CO_2 Sources

The carbon dioxide emissions from fossil fuel energy sources can be reduced substantially by means of cogeneration or combined heat and power (CHP) stations, which deliver both heat and electricity. This is because a higher proportion of the primary energy contained in the fuel at the power station is used compared to electricity-only power stations that do not make use of the waste heat. For example, in the U.K. the waste heat lost in power stations is comparable to the amount of heat provided by gas to heat buildings. Historically these CHP stations have been small power stations located in built-up areas that use heat mains (hot water pipes) to deliver heat to buildings within a certain radius of the CHP station. This approach has been successfully utilized in many European countries, particularly Denmark.

Small scale CHP is also employed for large buildings or small groups of buildings and can be successfully employed provided there is a good match with both the heat and the electricity demand (see Domestic Heating). There is also research underway in various locations (including Woking in the U.K.) into the use of fuel cells to provide electricity and heat from gas but with much reduced emissions. Fuel cells are electricity-generating devices that combine oxygen and hydrogen together, and the only important emission is water(see Fuel Cells). Research in the late twentieth century to develop fuel cells for vehicle propulsion could improve the economic prospects for fuel cell CHP and provide a means of utilizing emission-free hydrogen fuels derived from alternative or renewable energy sources of energy in the future. There are also field trials underway in the U.K. and in other European countries for so-

called micro-CHP units that can be used instead of gas boilers in buildings to produce both electricity and heat within the building. These micro-CHP units are mainly based around high efficiency Sterling engines, although some are based around fuel cells.

Obtaining Building Energy Supply from CO_2-Free and CO_2-Neutral Sources

Alternative or renewable energy sources such as solar, wind, or small-scale hydro energy and the low temperature heat stored in the ground are considered CO_2-free. Ground-coupled energy is primarily solar energy stored in the ground as low temperature heat; there may also be locations that are able to use geothermal heat from hot aquifers, which can provide hot water, and in some cases steam to power turbines to generate electricity. Most biological sources of energy (biofuels) are considered CO_2-neutral if they are derived from sustainable sources of biomass feedstocks. Building design can incorporate these energy sources in a variety of ways.

Solar energy:

- Trapping passive solar heat gains through windows or via solar sun-spaces.
- Incorporating active solar thermal collectors on roofs or façades to heat or preheat domestic hot water.
- Incorporating solar thermal collectors on roofs or façades of the building to actively trap solar energy for space heating via an intermediate insulated heat store.
- Incorporating photovoltaic (PV) modules on roofs or façades to generate electricity from light.
- Using high temperature solar collectors to power heat engines in climates that have high levels of direct sunlight.

Wind energy:

- Purchasing electricity from an electricity distribution company supplied by a wind power plant. Various "green electricity" rates or tariffs are available in many countries to assure their consumers that the electricity sold comes from a renewable energy source, frequently wind energy.
- Utilizing a small wind turbine located on a tower near to the building.
- Utilizing one or more community medium or large-scale wind turbines for group housing or neighborhood, village, or town. The town of Swaffham in the east of England obtains 70 percent of the annual electricity needs of its householders from two large wind turbines located nearby.
- Utilizing one or more medium or large scale wind turbines for educational, hospital, public, commercial, industrial, or agricultural buildings to offset electricity demand and generate an income from the export of excess electricity.
- Incorporating building integrated wind energy devices such as the *Aeolian Roof*™ or *Aeolian Tower*™ (and similar systems patented by the author) to generate electricity from wind passing over the building or be used to assist in the ventilation of the building.

Ground-coupled energy:

- Utilizing underground "earth-pipes" to temper incoming ventilation air.
- Utilizing a ground-source heat pump in combination with vertical or horizontal underground collectors to capture low temperature heat for space heating, hot water provision or cooling.
- Utilizing a local aquifer to provide "free cooling" of the building.

Small-scale hydroelectricity:

- Purchasing electricity from an electricity distribution company supplied by small-scale hydro power plant.
- Developing a small-scale hydroturbine for a building or group of buildings if a stream or river with sufficient fall and water flow is available locally.

Biofuels:

- Locally harvested wood fuels from sustainably managed woodlands can be used as a source of heat via wood pellet, wood chip, or log boilers when buildings are located in highly wooded or forested areas. (The supply must be within a minimum energy-balance distance from wooded areas if the wood fuel is transported by fossil-fueled transportation; otherwise the fuel consumption may exceed the energy content of the wood fuels being transported.) Both heat and electricity can be provided via cogeneration or CHP plant.
- In certain locations, agriculture crop residues may be used or converted into fuels and converted into electricity—for example,

the straw-powered generation stations in various European countries and at Ely in the U.K.—or used in a CHP plant.

- In certain locations energy crops may be grown and the fuels derived from them used in buildings if they are available locally.
- Gaseous fuels can be extracted from the anaerobic digestion of sewage, animal manure, crop wastes, food wastes, and certain components of municipal waste streams. The gas produced in this way can be used in buildings directly or via community or neighborhood cogeneration or CHP plants and district heating networks.
- It may be possible to recover energy from components of the waste stream by incineration, although this is a controversial approach due to concern about pollutants contained in their emissions, and it therefore requires very careful sorting of the waste to be used for combustion. It is an approach used in a number of European countries as waste-to-energy CHP plants.
- It is also possible to process components of the waste stream via pyrolysis (thermal decomposition in the absence of oxygen) to produce liquid or solid fuels, which can be used in conventional or CHP power stations.

See also **Air Conditioning; Domestic Heating; Electricity Generation and the Environment; Lighting Techniques; Power Generation, Recycling; Solar Power Generation; Wind Power Generation**

DEREK TAYLOR

Further Reading

Cofaigh, E., Olley, J.A. and Lewis, J. *The Climatic Dwelling: An Introduction To Climate-Responsive Residential Architecture*. James & James, London, 1996.

Crosbie, M.J. *Green Architecture: A Guide to Sustainable Design*. American Institute of Architects Press, Washington D.C., 1994.

Daniels, K. *The Technology of Ecological Building, Basic Principles and Measures, Examples and Ideas*. Birkhauser/Princeton Architectural Press, 1997.

Levine M.D., Martin N., Price L., Worrell E. *Energy Efficiency Improvement Utilising High Technology: An Assessment of Energy Use in Industry and Buildings*. World Energy Council, London, 1995.

Smil, V. *Energy in World History*, Westview, Boulder, CO, 1994.

Useful website

Department of Energy, Environmental Energy Technologies Division: http://eetd.lbl.gov/BT.html

Buildings, Prefabricated

Prefabricated buildings are assembled from components manufactured in factories. They differ in several ways from "stick-built" structures which are fabricated entirely on site. Typically, prefabricated components are mass produced out of the weather on indoor assembly lines. This method insures that parts can be replicated countless times with little or no variation. Economies of volume reduce costs, and precision measuring and cutting by stationary machine tools lessens waste. As work takes place on assembly lines, it is subject to constant inspection and quality control. Component assemblies made in immovable fixtures and forms further ensure that the finished work is precise and true. Thus, the quality of buildings made from parts fabricated on assembly lines has far greater chance of being accurate and uniform than those made in the field.

There are three basic types of factory made buildings and they differ according to the extent to which their components are assembled. Modular or sectional buildings consist of one or more finished rooms that leave the factory ready to be united on site with matching components. Panelized structures are produced as finished wall panels with both exterior and interior coverings in place or semifinished with only exterior cladding. At the building site, they are erected and joined with similar members to form walls. The precut building is delivered to the work site completely disassembled and consists of all the needed lumber cut to length. The latter is probably the most widely known type of prefabricated building.

One of the earliest examples of a prefabricated building was a house built in England in 1624. Its disassembled parts were shipped to Cape Ann, Massachusetts, where it was reerected. During the 1840s, buildings fabricated in New York, England, and as far away as China were shipped to California where they were used as housing for the influx of gold seekers. Prefabricated houses were being made in limited numbers by at least two U.S. factories by the 1890s.

In 1906, one of the first companies to offer mass-produced prefabricated housing was Alladin Redi-Cut Houses of Bay City, Michigan. Alladin kit houses consisted of precut and numbered pieces of lumber and they remained in production in various forms until 1981. Their renown and sales were eclipsed by Sears, Roebuck and Co., which over the years sold as many as 100,000 homes through its mail order catalog. House kits consisted of precut lumber, directions, and all the materials and

paint needed to complete the structure. As many as 450 designs were manufactured in Sears' factories between 1908 and 1940 when production ceased. Although they were neither trendsetting designs nor technologically innovative, they offered home buyers a choice of inexpensive products.

Generally, mass-produced buildings were the work of unnamed architects or designers. However, immediately after World War I, several well-known architects considered the potential of this building technique to meet Europe's pressing need for good and economical housing. A new type of architectural creativity was called for, and (Swiss-born) French architect Le Corbusier (aka: Charles Edouard Jeanneret) embraced the idea, developing several schemes for mass-produced prefabricated housing. However, despite his strong advocacy he produced little actual housing. Likewise, Walter Gropius of the Bauhaus in Germany, who had advocated industrialized worker housing as early as 1910, proposed structures composed of standardized flat roofed "Building Blocks," but little came of his proposals. During the mid-1930s, even noted America architect Frank Lloyd Wright, considered large scale, low-cost housing. While not actually advocating the mass production of housing, he realized the value of regularity or standardization of characteristics. This resulted in his design for small and relatively inexpensive Usonian Houses with standardized layouts, modular dimensions, and a modest degree of detailing. No more than 60 of these houses were built during a 15-year period beginning in 1936.

The sudden need for new housing just prior to and during World War II was met in part with prefabricated structures. One direct result of the demand was the development of the semicircular-shaped steel Quonset hut. That prefabricated structure was based on the British-built Nissen hut devised in the early 1930s. Construction methods as well as interior finish and insulation for the corrugated pressed steel Quonset were improvements over the Nissen hut. Quonsets were built in a number of sizes and served as housing for troops, warehouses, hospitals, and offices. Following the war, their use continued and broadened to include churches, shops, and emergency housing. Nonetheless, the need for domestic housing remained acute, and in the U. S. the situation was ideal for new approaches to prefabrication and mass production.

In 1947, New York developer Abraham Levitt and sons began to address the housing shortage on a scale previously unknown. In a Long Island potato field they commenced work on what would eventually become Levittown. Rather than focusing on building individual prefabricated houses, they set about building and selling an entire mass-produced community of standardized structures. The five simple types of homes they offered were built on concrete slabs. The homes were assembled of precut lumber shipped in from the Levitts's own lumbermill in California. There were great cost advantages to builder and buyer due to the tremendous volume of materials being used. Nails were needed in such quantity that the Levitts erected their own nail factory. Employing assembly line techniques, teams of workers performed specific tasks as they moved from one structure to the next and, by 1948, they were erecting 30 houses per day. When the development was finally completed in 1951, more than 17,000 homes had been constructed.

Carl Strandlund of Chicago established the Lustron Corporation and in 1948 began production of a new all-steel prefabricated house. Although the first prototype lightweight all-steel house was constructed by Albert Frey and A. Lawrence Kocher for a building products exhibition in New York in 1931, developer and consumer interest never materialized. However, with a pressing national need for housing immediately after the war, massive surplus industrial capacity along with generous federal government loans were available to address the problem. Assembly lines were set up in Ohio in a former aircraft factory manned by workers hired from the automotive industry. They stamped, punched, welded, and bolted sheet steel into house components. No wood was used in the construction, and exposed metal surfaces were preserved with a long-lasting porcelain enamel finish like that used on household appliances. Despite the suitability of the project to assembly line techniques and its acceptance by the public, production and financial difficulties put the Lustron Corporation in bankruptcy by 1950. Only 2500 prefabricated steel homes were built.

Nonetheless, in the decades following World War II, other all-metal prefabricated buildings were successfully manufactured and marketed and became a significant part of the industry. The Butler Manufacturing Company of Kansas City, Missouri was one of the first manufacturers of preengineered steel commercial structures. Devised by brothers Wilbur and Kenneth Larkin in 1939, their buildings used rigid steel frames. It was only after the war when restrictions on the sale and use of steel were lifted that Butler Buildings were first marketed. The rigid internal framework

followed wall and roof lines and required no internal supports or bracing. This left greater interior open space than in other buildings of comparable size. Precision manufacturing as well as quality control resulted in perfectly fitting components. Not only were the buildings readily assembled, but the economical use of materials insured a reasonable price. Butler produced a broad range of buildings for commercial and industrial users in an international market supplied by a network of manufacturing plants throughout the world. The company expanded the technology when it offered multistory and long-span preengineered steel buildings.

During the 1950s and 1960s the industry worked to dispel the belief, which began during World War II, that prefabricated buildings were shoddy and poorly made. One approach was to abandon the term prefabricated and refer to buildings as being either factory made, preengineered, or manufactured. Although the name changed, traditional assembly line prefabrication methods continued during the second half of the century. Standards were kept high and maintained by quality control and production oversight.

Prefabrication moved in a new, but not altogether permanent direction in the early 1970s, when architect Kisho Kurakawa unveiled plans for his Nakagin Capsule Tower in Tokyo. Designed and built between 1970 and 1972, it consisted of a steel reinforced concrete tower core filled with capsule living spaces. The modest size capsules were preassembled lightweight welded steel modules built and finished offsite. After delivery to the tower, they were hoisted into place and secured by four bolts. If necessary, the units could be detached and replaced.

At the close of the twentieth century, production of prefabricated buildings was strong, and the industry continued worldwide growth. Countries as disparate as Romania and Australia supported domestic manufacture, while in North America, Canada was a major international exporter of these buildings. In 1999, in the U.S. alone there were no less than 670 manufacturers who prefabricated wooden buildings and 579 who carried out the process in metal.

See also **Construction Equipment**

William E. Worthington, Jr.

Further Reading

Arieff, A. and Burkhart, B. *Pre Fab*. Gibbs Smith, Salt Lake City, 2002.

Carter, D.G. and Hinchcliff, K.H. *Family Housing*. Wiley, New York, 1949.

Rabb, J. and Rabb, B. *Good Shelter: A Guide to Mobile, Modular, and Pre Fabricated Houses, Including Domes*. Quadrangle/ Times Books, New York, 1975.

Stevenson, K.C. and Jandl, H.W. *Houses By Mail*. Preservation Press, Washington D.C., 1986.

Watkins, A.M. *The Complete Guide to Factory Made Houses*. E.P. Dutton, New York, 1980.

C

Calculators, Electronic

An electronic calculator is a calculating "machine" that uses electronic components, such as integrated circuits, transistors, and resistors to process the numbers that have been entered through a keyboard.

The electronic calculator is usually inexpensive and pocket-sized, using solar cells for its power and having a gray liquid crystal display (LCD) to show the numbers. Depending on the sophistication, the calculator might simply perform the basic mathematical functions (addition, subtraction, multiplication, division) or might include scientific functions (square, log, trig). For a slightly higher cost, the calculator will probably include programmable scientific and business functions.

Either way, the calculator is small, self-powered, and relatively disposable. Yet it was not always that way. The room-sized Z3 developed by Konrad Zuse and the ENIAC of the mid-twentieth century are considered the world's first digital computers (see Computers, Early Digital), but they really were little more than electronic calculators designed to work through a series of numbers. The fact that these early behemoths had the computing power of today's $10 student calculator does not lessen their importance to the engineers and scientists of the mid-1940s. This was a time when people would spend days, weeks, or months calculating formulae using mechanical adding machines. Or they would use slide rules and preprinted tables to help determine certain scientific solutions, knowing that the final number was close but not perfectly accurate to the *n*th degree.

The need for sophisticated and speedy manipulation of numbers was of major importance. It took the birth and growth of the electronic age for this to happen. Vacuum tubes (valves) allowed the beginning of the computing age, transistors brought it down to size, and integrated circuits were the catalyst to make electronic calculators truly possible and accessible to the world. All this happened in the space of thirty years.

In 1962 British company Sumlock Comptometer designed and sold the first all-electronic calculator, a desktop model named the Anita (an acronym for A New Inspiration To Arithmetic, as the story goes). This model used small vacuum tubes as electronic switches. The reliability was not infallible, but it was much faster than the electromechanical calculators available at the time and very desirable to people who needed the number-crunching "computing" power (see Figure 1).

Figure 1. Anita Mk VIII, one of the world's first desktop calculators, launched in 1961.
[*Courtesy of www.vintagecalculators.com.*]

In the early 1960s, Japanese manufacturers were among the first to become interested in electronic calculators and the first to develop all-transistor devices. Sony showed a prototype at the New York World's Fair in 1964, and Sharp was the first to sell a production model soon after. Even though it cost the same as a small car ($2500) at the time, Sharp's CS-10A sold well to those who needed its speed and power. However, at 25 kg it was not portable.

It was the invention of the integrated circuit, the heart of modern-day computers, that made the most serious mark in electronic calculators. An integrated circuit can reduce the size of the circuitry in an electronic device to one hundredth (or more!) of the original size, and offer computing power far greater than was otherwise possible.

In the late 1960s, calculator companies, particularly Japanese, began to work with American semiconductor companies who were developing the ability to design and manufacture integrated circuits.

In 1967 one such company, Texas Instruments, had a breakthrough. Their engineers took an electronic calculator design with hundreds of discrete components (transistors, diodes, relays, capacitors) that covered the top of a desk and successfully reduced most of the electronics to four small integrated circuits. Interestingly enough the project, code-named Cal-Tech, was created not to start Texas Instruments' entry into calculators, but to interest calculator makers in using their integrated circuit products. The integrated circuit was then a new and unproven product in the world of electronics, and the Cal-Tech was intended to prove that they were effective. Texas Instruments' effort worked—Canon became very interested in the Cal-Tech and eventually created a similarly designed calculator (the 1970 Pocketronic) with the help of Texas Instruments' engineers. In the same period, other business machine companies began to develop and market similar products.

One small Japanese company, Busicom, began to work with a fledgling company called Intel to develop an integrated circuit to use as the brains for Busicom's 141-PF desktop calculator. Eventually, Intel bought back the rights to the circuit design. This design, the model 4004 microprocessor, is the grand predecessor to all of Intel's current Pentium products.

Companies like Bowmar and Summit in the U.S., and Sinclair and Aristo in Europe, would develop very small pocket-sized models in the early 1970s. Soon after, hundreds of large and small companies worldwide would develop and sell hand-held calculators, thanks to the availability of inexpensive, calculator-function integrated circuit chips.

A major milestone occurred in January 1972 when Hewlett Packard (HP) sold the world's first scientific pocket calculator. This model was in so much demand that even though it cost US$395 (two weeks wages for most engineers) there was a 6-month backlog to buy one. The HP-35 was so powerful that it rivaled some small computers and brought computing power directly to the hands of its users. As the years passed in the 1970s and 1980s, the production cost of integrated circuits (once costing US$100 each) dropped to less than $1. Calculators which once cost US$300 dropped to $10 or less. By that time, simple models were even distributed free as inexpensive promotional tools.

At the end of the twentieth century, the electronic calculator was as commonplace as a screwdriver and helped people deal with all types of mathematics on an everyday basis. Its birth and growth were early steps on the road to today's world of computing.

See also **Calculators, Mechanical and Electromechanical; Computers, Early Digital; Integrated Circuits, Design and Use; Transistors**

GUY D. BALL

Further Reading

Ball, G. and Flamm, B. *Collector's Guide to Pocket Calculators*. Wilson/Barnett, Tustin, CA, 1997.
Greenia, M.W. *History of Computing: An Encyclopedia of the People and Machines That Made Computer History*. Lexikon Services, Antelope, CA, 2000.
Haddock, T. *Personal Computers and Pocket Calculators*. Nostalgia Publishing, La Centre, 1993.
How the Computer Got Into Your Pocket. *Invention and Technology*, Spring 2000.
Johnstone, B. *We Were Burning: Japanese Entrepreneurs and the Forging of the Electronic Age*. Basic Books, New York, 1999.
Lukoff, H. *From Dits to Bits: A Personal History of the Electronic Computer*. Robotic Press, Ontario, 1979.
Reed, T.R. *The Chip: How Two Americans Invented the Microchip and Launched a Revolution*. Random House, New York, 2001.
Various. Special 15th Anniversary Issue. *Electronic Engineering Times*, November 16, 1987.

Calculators, Mechanical and Electromechanical

The widespread use of calculating devices in the twentieth century is intimately linked to the rise of large corporations and to the increasing role of mathematical calculation in science and engineer-

ing. In the business setting, calculators were used to efficiently process financial information. In science and engineering, calculators speeded up routine calculations.

At the beginning of the nineteenth century, mechanical calculators were already in widespread use, and by 1822 Charles Babbage was at work on his difference engine (which he never completed because of the mechanical complexity of thousands of brass cogs and gears). Based on technology developed over several centuries, early twentieth century calculators can be divided into two major types: (1) the slide rule; and (2) the adding machine and related devices.

Invented by William Oughtred in the seventeenth century, the slide rule is based on the physical relationship between two logarithmic scales. In the mid-nineteenth century, accurate methods for reproducing such scales on instruments were developed, and the slide rule became much more widely available. By the early twentieth century, slide rules were commonly used by scientists and engineers for calculations involving multiplication, division, and square roots.

During the first half of the twentieth century, slide rules underwent a gradual evolution. Plastic gradually replaced wood and aluminum in construction, and additional scales were added to more expensive slide rules for specialized computation. In addition, a variety of specialized slide rules were developed for particular applications, most notably electrical engineering.

Slide rules had several advantages over manual calculation. First, with practice, calculations could be made much more quickly than by hand. Second, slide rules were compact and light, and so could be used almost anywhere. Finally, calculations could also be carried out with a reasonable degree of accuracy. However, slide rules also had one important disadvantage—the results of their calculations were approximations rather than exact numbers. This was acceptable for most scientific and engineering calculations, but unacceptable for financial calculations, which had to be exact. As a result, the use of slide rules was limited primarily to scientific and engineering applications.

The second major type of mechanical calculator was the adding machine and related devices. While a variety of adding machines were developed beginning in the seventeenth century, a series of major innovations in the late nineteenth century led to the widespread marketing of machines that were much more reliable, compact, and easy to use. The combination of increased demand from large business and government organizations, along with improvements in machine tools, materials, and manufacturing techniques, triggered a period of intense innovation that continued until World War I. The key mechanical innovations were all patented before 1900, but putting them into practice took some time, and it was not until the first years of the twentieth century that mechanical calculators began to sell in large numbers (see Computers, Uses and Consequences).

Two key mechanical innovations just before 1900 made this expansion possible. The first, invented almost simultaneously by the American Frank S. Baldwin and the Swede Willgodt Theophil Odhner, consisted of a round disk with moveable radial pins that could be extended beyond the edge of the disk. Input for calculation was a function of varying the number of pins extended by the action of levers, which then meshed with a register mechanism. This design, known both as Baldwin type and Odhner type, proved much more compact and reliable than previous systems.

The second was the introduction of the keyboard for data entry. The first practical machine to use a keyboard for data entry, the Compometer, was invented in 1885 by the American Dorr Eugene Felt, who formed the Felt and Tarrant Manufacturing Company in 1887 to produce his device. The keyboard entry system greatly simplified the operation of calculating machines, and was soon copied by other manufacturers, such as William Seward Burroughs, who founded the American Arithmometer Company (which became the Burroughs Adding Machine Company in 1905). These machines had nine rows of keys, one for each digit (1 to 9). The number was entered by pressing one digit in each column. There was no zero key because zero was represented by the absence of a keystroke in the corresponding column.

During the first half of the twentieth century, the market for mechanical calculators was divided into roughly three categories of machines. The first, and largest in terms of total numbers produced, was adding machines. These machines, produced in large numbers in standard designs that changed little over time, were used primarily for basic accounting by small businesses. Their manufacture was dominated by three large firms: the Burroughs Adding Machine Company and Felt and Tarrant in the U.S. and Brunsviga in Germany, though a variety of smaller firms also competed.

The second type of mechanical calculator was the four-function calculator, which were similar mechanically to adding machines, but performed

multiplication and division in addition to addition and subtraction. Such calculations could be carried out on adding machines by trained operators, but four-function calculators were faster and allowed the use of less highly trained operators. Four function machines were used primarily in medium-sized businesses whose volume of calculation did not justify the use of more expensive specialized machines. They were also used by engineers, which led to the development of machines in the 1950s that included specialized functions such as calculating square roots, such as the Friden SRW model of 1952 (which weighed 19 kilograms). No one firm dominated the market for these machines, and there was considerable competition for market share.

The third type of mechanical calculator was a group of devices known as accounting machines (also called book-keeping machines during the 1920s and 1930s). These devices were used to enter data onto standard forms and then to perform accounting calculations. They could also prepare balances and print the results. Used by larger firms whose volume of calculation could justify the investment in specialized machinery, in some applications accounting machines competed directly with punched card tabulating systems, such as those developed by Herman Hollerith in 1890 and later developed by the successor to Hollerith's company, International Business Machines (IBM). Accounting machines were also used in specialized niche applications.

All three types of calculator could be either hand or motor driven in operation. Motor-driven mechanical models first appeared just after 1900. By replacing the hand crank with a small electric motor, these machines were less tiring to operate and could reliably perform repeat operations, simplifying the construction and operation of machines that performed multiplication and subtraction. As a result, motor-driven calculators were commonplace on the desks of engineers by the 1940s. Hand-driven machines continued to sell well, however, because they were quieter, lighter, smaller, and less expensive.

Regardless of the type of mechanical calculator, firms found that to be successful they had to provide a high level of service to customers. Calculators were sold, not bought, and firms maintained large sales forces to educate customers as to the capabilities of machines and to anticipate customer's needs. Calculator manufacturers also had to service their machines, which as mechanical devices need constant maintenance to function reliably, and train operators in the correct operation procedures.

The manufacture and sale of calculators was a widespread industry, with major firms in most industrialized nations. However, the manufacture of mechanical calculators declined very rapidly in the 1970s with the introduction of electronic calculators, and firms either diversified into other product lines or went out of business. By the end of the twentieth century, slide rules, adding machines, and other mechanical calculators were no longer being manufactured.

See also **Calculators, Electronic; Computers, Analog; Computers, Uses and Consequences**

Mark Clark

Further Reading

Aspray, W., Ed. *Computing Before Computers*. Iowa State University Press, Iowa, 1990.

Cortada, J.W. *Before the Computer: IBM, NCR, Burroughs and Remington Rand and the Industry They Created, 1865–1956*. Princeton University Press, Princeton, 1993.

Hopp, P.M. *Slide Rules: Their History, Models, and Makers*. Astragal Press, 1999.

Jerierski, D.V. *Slide Rules: A Journey Through Three Centuries*. Astragal Press, 2000.

Marguin, J. *Histoire des instruments et machines à calculer: trois siècles de mécanique pensante, 1642–1942*. Harmann, Paris, 1995.

Martin, E. *The Calculating Machines (Die Rechenmaschinen): Their History and Development (1925)*. Translated and edited by Kidwell, P.A. and Williams, M.R. MIT Press, Cambridge, MA, 1992.

Turck, J.A.V. *Origin of Modern Calculating Machines (1921)*. Arno Press, 1972.

Cameras, 35 mm

The origins of the 35 millimeter (mm) camera lie in the increasing availability of motion picture film stock during the early years of the twentieth century. The 35 mm format was first used in Edison's Kinetoscope, a moving picture viewing device patented in 1891, and was later adopted as the standard film gauge by cinematographers after 1896. The earliest 35 mm film was very slow by the standards of the day and not ideal for still camera work. However, as the quality of the film improved, the potential virtues of small size and convenience of handling began to appeal to camera designers.

Although three Spanish inventors took out a British patent for a still camera using 35 mm film as early as 1908, the first 35 mm still camera sold to the public was probably the American Tourist Multiple camera of 1913. Patented a year later, the design was for a camera taking a 15-meter magazine of standard 35 mm cine film, allowing 750 still exposures. Another American 35 mm camera, the

Simplex, appeared in 1914. The Simplex camera allowed the photographer to switch between two picture sizes, giving 400 full-frame pictures or 800 half-frames, or any number in-between if the sizes were mixed. The Jules Richard Homeos camera, a stereoscopic 35 mm camera, was patented in France and England in 1913 but did not reach the market until a year later.

Sales of the cameras above were limited by the outbreak of World War I in 1914, as was the development and marketing of further models. Unsurprisingly perhaps, the biggest selling 35 mm camera of the decade was produced in America, where public spending was least affected by war. However, the Kodak 00 Cartridge Premo box camera of 1916 was an unusual model using unperforated 35 mm film, a concept that was to have little influence on future development. More important was the work of Oskar Barnack, an engineer working with cine cameras for the Leitz Optische Werke in Germany. In 1913 Barnack designed and built for his personal use a small still camera to make use of "short ends" of cine film. War prevented immediate development but Barnack's little camera was the prototype of the Leica, a camera that was to have a profound influence on future camera design.

The original Leica is sometimes thought of as the first 35 mm camera but it was far from that. During the early 1920s several new 35 mm models were marketed throughout Europe, some enjoying modest success. Nevertheless, the most significant camera of the period was undoubtedly a developed version of Barnack's 1913 camera, the Leica of 1925. Manufactured by Leitz, a renowned optical instrument company, its most important features were the compact integrated design and high-precision construction along with the quality of its lenses, which allowed fine prints to be produced from tiny negatives. The Leica convinced the photographic world that the 35 mm camera was worthy of consideration by the serious photographer. Improved models included provision for a screw fitting interchangeable lens system, which partly promoted another important innovation, the coupled rangefinder. The first coupled rangefinder 35 mm camera was the Leica II introduced in February 1932. A month later the giant German photographic company, Zeiss Ikon, produced its own precision constructed 35 mm camera, the Contax. From the outset, the Contax came with a coupled rangefinder and a bayonet fitted interchangeable lens system. Both cameras had their own reloadable film cassettes, a system that was to lead to universal cassette standardization.

More manufacturers began to exploit the 35 mm format. A noteworthy example was the 1934 Retina camera, the first in a series of well-made 35 mm self-erecting folding cameras marketed by the German arm of Eastman Kodak. The Retina models proved popular and provided a comparatively inexpensive alternative to the Leica and Contax for the keen amateur photographer. The first orthodox mass-produced 35 mm camera to be produced in America was the Argus model A camera of 1936. It was an immediate success as were subsequent models, one of which (the CC model) had an integral photoelectric exposure meter, the first American camera to be so equipped. The Argus C-range cameras, widely known as "bricks" in the trade, were particularly popular. Other significant cameras of the period include the German Robot camera of 1934, the first purpose-designed 35 mm camera with a spring-driven motor drive, and two 35 mm single-lens reflex cameras, the Russian manufactured Sport of 1935 and, more influentially, the German Kine Exakta of 1936. The latter camera was a particularly important pointer to the future.

Coupled rangefinder cameras continued as the classic 35 mm design throughout the 1940s. The postwar Leica remained essentially similar to its original form until a radical redesign led to the introduction of the M series cameras in 1954. The Contax also underwent little major development. Both however, continued to significantly influence camera manufacturers, particularly in Japan. In the immediate postwar years Canon produced a series of Leica inspired rangefinder cameras while the Nikon S of 1954 was firmly based on the Contax II. This period also saw the production of enormous numbers of cheaper 35 mm cameras of varying quality.

Good-quality 35 mm rangefinder cameras from both Europe and Japan remained popular for most of the 1950s. It was the end of the decade before single-lens reflex cameras from Japan began to seriously challenge their supremacy. Yet only a few more years passed before reflex cameras from the likes of Pentax, Nikon, and Canon, incorporating important innovations such as the pentaprism viewfinder, instant-return mirror and through-lens metering had almost completely displaced rangefinder cameras from the quality 35 mm market. A limited number of fixed lens nonreflex 35 mm cameras with excellent specifications continued to enjoy good sales, as did many so-called "automatic" snapshot cameras. Nevertheless, there was a steady decline in the use of the 35 mm photographic system during the last third of the twentieth

century. Its main representative today is the high-quality, pentaprism single-lens reflex camera.

See also **Cameras, Automatic; Cameras, Single Lens Reflex; Cameras: Lens Designs, Wide Angle, Zoom**

John Ward

Further Reading

Coe, B. *Cameras: From Daguerreotypes to Instant Pictures*. Marshall Cavendish Editions, London, 1978.

Coe, B. *Kodak Cameras: The First Hundred Years*. Hove Foto Books, Hove, 1988.

Condax, P.L, Tano, M.; Takashi, H.; Fujimura, W.S. *The Evolution of the Japanese Camera*. International Museum of Photography, Eastman House, Rochester, 1984.

Dechart, P. *Canon Rangefinder Cameras 1933–68*. Hove Foto Books, Hove, 1985.

Hicks, R. *A History of the 35 mm Still Camera*. Focal Press, London, 1984.

Janda, J. *Camera Obscuras. Photographic Cameras 1840–1940*. National Museum of Technology, Prague, 1982.

Miniature Camera in *The Focal Encyclopedia of Photography*, vol. 2, fully revised edn. London, Focal Press Ltd, London, 1965.

Rogliatti, G. *Leica The First 60 Years*. Hove Foto Books, Hove, 1985.

Rotoloni, R. *Nikon Rangefinder Camera*. Hove Foto Books, Hove, 1983.

Tubbs, D.B. *Zeiss Ikon Cameras*. Hove Foto Books, Hove, 1977.

Cameras, Automatic

The quality and precision of a photograph is basically dependent on the aperture, the light sensitivity of the film, and the shutter speed of the camera. As photography became a sought-after source of record keeping for the amateur, demand increased for a camera that could automatically align this seemingly nefarious collaboration of technical details. The evolution of the automatic camera inevitably reflects the mechanical, electronic, and high-tech progress of the twentieth century, but primarily employed its own specific mechanical innovations.

Prior to the twentieth century, cameras were a fairly cumbersome collection of enormous photographic plates and large boxes, usually mounted on massive tripods in order to keep camera perfectly steady for an image to burn the film without blurring. In the 1870s faster-exposure plates were introduced and the shutters that controlled light exposure were built for the first time inside the camera, reducing some of the bulk. By the turn of the century a wide variety of cameras were available for every special purpose imaginable, including gigantic cameras that produced large prints and double-lensed "stereo" cameras that created three-dimensional and panoramic images. But the amateur photographer was left out of these innovations, too inconvenienced to master the complicated equipment and heavy, expensive photographic plates.

In 1888, American George Eastman brought "Kodak No. 1" to market, as the first hand-held camera, which used rolls of film in place of the photographic plate. While this "snapshot" camera was quite popular, quality photographs from a hand-held camera that could manipulate the necessary components—aperture, shutter speed, and film speed—were decades away. The Kodak No. 1 was set at a single speed shutter of 1/25 of a second, and it used a fixed-focus lens good for any subject more than 2.5 meters away. The camera came preloaded with a roll of 100 pictures, which the photographer had to send, along with the camera, back to Kodak to have developed and reloaded.

It wasn't until 1925, at the Leipzig Fair, that German E. Leitz introduced the Leica camera, which used a higher-quality 35 mm film, a fast focal-plane shutter, and a more precise f/3.5 lens. These adaptations created a camera that could take high-quality photographs under a variety of conditions. The focal plane shutter could give a range of exposures, from 1/1000 second to 1 second, and the lever that operated this mechanism also advanced the film. The lens, which was fast enough to allow for indoor photography, was also connected to a rangefinder that focused easily at the hand of the photographer. The Leica ushered in a new era of photographic conveniences, and soon photojournalism became a fixed—and critical—customer of these new hand-held cameras.

Soon thereafter, the growing needs of photojournalists and discriminating amateur photographers inspired the creation of similar handheld cameras with slight improvements over the Leica. In 1927, Rolleiflex introduced the twin-lens reflex camera, employing a separate focus lens through which the photographer looked, and a parallel lens that captured the image. Ten years later, a single-lens reflex (SLR) hand-held called the Exacta came to market, which allowed the photographer to easily focus through one 35 mm lens. The shutter of the first SLR cameras was dependent on a fast-moving slit that exposed different parts of the subject at different times across the film, distorting fast-moving objects. But the technology of the SLR cameras eliminated the parallax error—or the difference between the image that goes through the lens and the image seen through the view-

finder—of the smaller viewfinder cameras. Using a mirror and a diaphragm, the SLR camera allows the photographer to both compose the image and focus through the 35 mm lens.

Following these improvements, a series of cameras that appealed to the novice photographer came on the market. Most notably, American Edwin H. Land created a camera that used a film capable of producing developed shots on demand—the precursor to the Polaroid camera. After World War II, a sudden plethora of camera manufacturers produced cameras that used 110 film and electronic controls, taking Eastman's snapshot Kodak No. 1 a step further by allowing a slight range of shutter speed, but still employing the same basic fixed-focus lens.

Lens and optical companies soon designed their own versions of the Kodak and Leica amateur-friendly cameras. By 1946, Japanese companies developed 35 mm cameras. All these cameras still required imprecise threading of film, manual control over focus, aperture and shutter speed. The next step in improvements came in the form of film rolls that snapped into a fitted compartment and threaded themselves. Eventually, automatic control over exposure was also incorporated, using a built-in light meter that reads the level of illumination, then sets either the aperture, or shutter speed, or both together. In general, these two forms—the leaf shutter and the focal-plane shutter—are still used today. Originally in Leicas's historical camera, the leaf shutter consists of a series of overlapping blades in the lens, powered by a spring-loaded shutter button. When the shutter is pressed, the blades open and shut according to the settings. The focal-plane shutter, on the other hand, is not in the lens but the camera itself, directly in front of the film. It is a faster mechanism, allowing for faster exposure, utilizing two curtains that open slightly—also by a spring-driven shutter button—exposing the film to a window of light.

Automation of shutter speed and aperture accompanied the development of the built-in exposure meter. This is done by "reading" the light reflected from the object being photographed by a needle in the light-measuring device. The needle measures the light, moving into position. When the shutter button is released, the needle becomes fixed, and a scanner arm moves until it hits the needle. The end of the scanner arm then engages the devices that control the shutter and the aperture diaphragm.

By the late 1950s, camera companies included extras like timers, and by the late 1960s, Polaroid created the first color instant film and the first "instamatic" camera, which was similar to the 110-film cameras of the forties, but with color film.

Autofocus, present in the 1950s in its most basic form, utilized a motor that spun the lens's focus ring, which was usually inside the lens. Over time, manufacturers built the motor in the body of the camera, but the poor quality of these early systems earned autofocus a bad reputation (especially among expert photographers). By the 1970s, however, the American-company Honeywell earned a patent for phase-detection autofocus, which was applied to Konica's C35 AF camera. Basically, the system works by taking light from the subject that passes through the lens and the semitransparent reflex mirror. Another mirror directs the light toward the autofocus module, and an array of light-sensitive charge-coupled devices, or CCDs. The distance between the CCDs determines the focality of the image. A small circuit board then controls a motor that moves the focusing ring of the lens.

By the 1980s, compact, "pocket-sized" 35 mm cameras with autofocus came onto the market, employing systems developed mostly by Japanese manufacturers, who fine-tuned the Honeywell technology and added computer technology and microelectronics to the mechanism controlling the systems. In addition, the perils of setting aperture and shutter speed were further relieved by a new kind of light meter employing a panel of semiconductor sensors that translate the light level into electrical energy. The light meter then "reads" the film speed by special markings on the outside of the 35 mm cartridge and accounts for shutter speed to adjust for the correct aperture.

By the early 1990s, both automatic focus and light meters required nothing of the shooter. Central microprocessors, now fairly common in cameras, activate several motors that control focus, shutter speed and aperture, in a single click. Automatic cameras at the beginning of the 21st century are closer to computers, especially as digital cameras have become a standard piece of equipment for the amateur.

See also **Cameras, 35 mm; Cameras, Polaroid; Cameras, Single Lens Reflex**

LOLLY MERRELL

Further Reading

Auer, M. *The Illustrated History of the Camera from 1839 to the Present*. New York Graphic Society, Boston, 1975.

Dobbs, W.E. *Your Camera and How it Works*. Ziff-Davis, Chicago, New York, 1939.

Horne, D.F. *Lens Mechanism Technology*. Adam Hilger, London, 1975.
Jenkins, R. *Images and Enterprise: Technology and the American Photographic Industry, 1839 to 1925*. Johns Hopkins Studies in the History of Technology, Maryland.
Johnson, W.S. *Photography from 1839 to Today*. Klotz, George Eastman House, Rochester, 1999.
Ray, S.F. *Camera Systems: A Technical Guide to Cameras and Their Accessories*. Focal Press, London, 1983.
Tubbs, D.B. *Daguerre to Disc: The Evolution of the Camera*. California Museum of Photography, Riverside, 1983.

Cameras, Digital

Digital photography constitutes the most revolutionary development in photography since the early experiments of image making on chemically sensitized materials in the 1820s, just as digital computing technologies revolutionized the manner in which the written word was recorded in the latter half of the twentieth century. The invention of flexible film in the late nineteenth century profoundly changed the course of photography, but the adoption of electronic media 100 years later as a substitute for film constitutes an even more fundamental upheaval, with broad implications for the future of visual communication. At the end of the twentieth century electronic imaging seemed poised to render chemically processed photographs obsolete, and the future of film appeared to be in doubt as improvements in the quality and capacity of electronic imaging occurred at a dizzying rate.

Digital photography blends three distinct technologies: traditional photographic methodology, video, and digital computer systems. The advent of television and the cathode-ray tube provided both the technology and the conceptual framework for the production of images on an electronic screen instead of film or paper. The digital computer offered the means for rendering not only text but pictures, broken into tiny picture elements called pixels, which could approximate continuous-tone pictorial representations if the number of constituent pixels was large enough and their size small enough to resist detection by the eye.

Video and computer technologies are electronic cousins, although the first uses analog means and the latter digital. A half century's experience in viewing images on a television screen provided a conceptual basis for the easy reception and popularization of digitally produced images on a computer monitor, and the notion of transmitting televised imagery in the air has its parallel in the transmission of digital imagery via the Internet. Very quickly the concept of imprinting digital images on paper would threaten the traditional chemically based photographic print.

The computer scanner initially served as an intermediate tool between traditional photography and the digital camera. The rise of the Internet created a demand for transmitting photographic imagery, and scanners converted photographs into electronic files which could be viewed on a computer monitor and shared, either by recording them on a medium such as a floppy disk for exchange with other computer users, or by transmitting them via various file-sharing protocols, including the familiar e-mail systems in common use. Computer printers could convert electronic image files into prints, which eventually rivaled the ordinary chemically based photographic print. The rise of electronic image editing and manipulation programs, such as Adobe Photoshop, served to aid in the preparation of scanned images. At first digital photography was largely confined to this conversion of film-based photographs into electronic copies, but soon a new generation of cameras emerged to produce direct electronic images without an intermediate chemically processed photograph at all. This method, interacting directly with a computer, is especially attractive because it completely eliminates the time-consuming steps of chemical processing and the scanning.

There are two types of electronic still cameras, the still video camera and the digital. Both are similar to film cameras except that a built-in light-sensitive computer chip or "imaging array," either a charge-coupled device (CCD) or a complementary metal-oxide semiconductor (CMOS), takes the place of film to capture images. The electronic era in still photography essentially began in 1981 with the introduction of Sony's Mavica video camera, but still video was an analog system. The Dycam digital camera initiated the fully digital era in 1990.

When light enters the digital camera through the lens and strikes the chip, it emits an electrical charge which is measured electronically and is then sent to the electronic memory (buffer) of the camera. It is then compressed into a format; for example, a JPEG, and transferred to the memory card or disk. Some cameras need to wait for the completion of this process before another picture can be taken, while others have a buffer large enough to hold several pictures, thus enabling rapid "burst" shooting for a group of consecutive pictures. The CCD has millions of receptors to record the amount of light striking them. Each sensor represents a pixel or image element. The information on the chip is read one horizontal line at a time into the internal memory of the camera, combining the individual

pixels into an image, before it is saved to the memory medium of the camera, from which it can be viewed or printed via a computer.

Color can be recorded with a digital camera in several ways. Some cameras use three separate sensors with a separate filter, and light is directed to the sensors by a beam splitter. Another method is to rotate a series of red, green, and blue filters in front of a sensor, but the continuous movement of the filter wheel requires stationary subjects. Another sophisticated idea places a permanent filter over each sensor and utilizes an interpolation process to approximate color patterns. Most consumer cameras use a single sensor with alternating rows of green–red and green–blue filters in what is called a Bayer filter pattern.

The quality of a digital photograph depends upon a complex combination of factors. The higher the resolution, the more closely the image resembles a continuous-tone analog image. Resolution in film-based photography relates to both the ability of the camera lens and the recording medium to reproduce or ''resolve'' fineness of detail; in digital photography the number and size of the pixels that constitute the image, control resolution. The larger the number and the smaller the size of the pixels, the greater the resolution. Since higher resolution results in greater file size, practical considerations usually mandate a compromise on resolution. Low resolution can be satisfactory for viewing images on a computer monitor, but images to be printed normally require a higher resolution. Most digital cameras permit the photographer to select a ''capture'' resolution, which can later be modified with the computer and the editing program, but the photographer must also consider the amount of camera system memory the image will consume. If a preferred resolution setting will utilize too much memory, it may be necessary to select a higher image compression to reduce the file size. The need to manipulate and balance these factors to arrive at acceptable results demonstrates the inherent limitations of digital photography and is one reason that film photography still maintained a strong presence at the end of the twentieth century. Yet the rapid pace of technological advances in computers, software, and digital cameras has led many to assume that film-based photography's days are numbered.

The convenience of digital media has led many professional and other serious photographers to abandon film-based photography altogether. As the tempo of modern life has accelerated, most newspaper photographers and other photojournalists have embraced digital photography exclusively. ''Wet'' darkrooms were dismantled because photographers in the field, using expensive, sophisticated digital cameras, can transmit images electronically and get them into print in a fraction of the time required with film cameras; also, chemical photography is considered environmentally unfriendly. Photographic manufacturers such as Eastman Kodak, Fujifilm, Canon, Pentax, Olympus, Leitz, and Nikon devote an increasingly higher percentage of their product lines to digital cameras, joined by such familiar electronic firms as Sony, Samsung, and Panasonic.

With the major exception of the image recording system, film and many digital cameras have similar design principles, employing a range of shutter speeds and aperture sizes to control exposure. Some digital cameras employ an entirely different exposure system in which the chips in the image-sensor array turn on and off at varying intervals in order to capture more or less light. Other digital cameras vary exposure by varying the strength of the electrical charge that a chip emits in proportion to the amount of light received. Despite the radical difference of this method, the camera's sensitivity or ''speed,'' which may be manually variable, is typically rated according to standard film speed nomenclature. The resolution rating of a camera generally is expressed in megapixels, ranging from one to four and higher. The higher the megapixel rating of the camera, the more expensive it tends to be. The larger the pixel count, the larger an acceptable print that can be made, but the higher the camera price and the more file storage space an image requires.

Early digital cameras used internal or ''onboard'' memory for file storage, requiring the periodic transfer of images to a computer after the memory was filled before additional pictures could be taken. This limitation was solved by the advent of removable memory devices, such as cards or disks. Most later digital cameras employed a liquid crystal display screen to preview pictures, which functioned as a viewfinder, as well as to review images stored in the camera, as with a computer monitor; this feature represents a substantial advantage over film cameras, permitting the user to redo an unsatisfactory image before leaving the scene. A variety of innovative special-purpose and special-feature digital cameras were also marketed.

DAVID HABERSTICH

Further Reading

Aaland, M. and Burger, R. *Digital Photography*. Random House, New York, 1992.

King, J.A. *Digital Photography for Dummies*, 4th edn. Wiley, New York, 2002.
B&H. Photo-Video, Inc. *Digital Photography Source Book*, 3rd edn. B&H, New York, 2003.
Kasai, A. and Sparkman, R. *Essentials of Digital Photography*. Translated by Hurley, ENew Riders, Indianapolis, 1997.
Ritchin, F. *In Our Own Image: The Coming Revolution in Photography: How Computer Technology Is Changing Our View of the World*. Aperture, New York, 1990.

Useful Websites

Nice, K. and Gurevich, G.J. *How Digital Cameras Work*: http://www.howstuffworks.com/digital-camera.htm

Cameras, Disposable

The first disposable camera was probably "The Ready Fotografer" of 1892, which used a pinhole aperture instead of a lens. But the single-use concept lay dormant for decades. At the end of the twentieth century digital photography threatened to make photographic film obsolete, but one factor forestalling this revolution was the popularity of "disposable" 35 mm rollfilm cameras. They recapitulate several themes of technological and socioeconomic significance for the history of amateur photography from the late nineteenth century through the twentieth century.

The first theme is epitomized by the Eastman Kodak slogan, "You press the button, we do the rest," for the Kodak camera of 1888 and its successors. The Kodak was a simple box camera, preloaded with rollfilm for 100 exposures, intended for an inchoate amateur market. When the customer finished the roll, the entire camera was sent to the company for processing and printing, and the reloaded camera was returned to the owner. Thus one could take pictures without handling film at all, let alone indulging in messy darkroom work. The primary innovation that attracted consumers was the processing service, since the darkroom was the domain of professionals and specialists. Rollfilm also eliminated the need to handle individual glass plates. The amateur market for cameras and film was created virtually at a single stroke. The other advantage of the system—freedom from loading and unloading film—was forgotten when later cameras accustomed users to performing these tasks. Perhaps having a camera on hand at all times seemed preferable to waiting days for a reloaded Kodak to return.

Another advantage of the Kodak was its small size compared to the large, tripod-mounted cameras of professionals. It was hand-held in operation and light to carry. Later, cameras became even more portable; for example, by using a bellows so that the camera could be collapsed when not in use. Folding cameras were called "pocket" cameras, and Kodak appropriated the word as a brand name, although one needed large pockets to accommodate them. The trend toward smaller cameras was well established by the early twentieth century when the 1924 Leica appeared. This precision camera used 35 mm film, was eminently suitable for a pocket, and its superior optics made enlargements from its small negatives feasible. The first Leicas were difficult to load, however, and that issue plus the high price made it the choice of professionals and dedicated amateurs, not casual snapshooters. 35 mm cameras with wide-aperture lenses and adjustable shutter speeds were versatile but expensive and were considered "serious" cameras. Meanwhile Kodak and other manufacturers continually introduced new fixed-focus, inexpensive cameras for amateurs who could not afford adjustable cameras or preferred simplicity. At first, these simple cameras were limited to exposures in bright sunlight, but eventually battery-operated flashbulb guns were added to camera kits.

In the 1970s, many inexpensive cameras used 110 format film, whose frame size was about half that of 35 mm. The film came preloaded in plastic cartridges that could drop into the camera like an audio cassette. Some were essentially novelty semidisposable cameras, such as a lens assembly attached to a keychain, which snapped onto a 110 film cartridge. Frequently disdained, 110 cameras nevertheless had features that anticipated aspects of 35 mm disposable cameras. When the latter half of the twentieth century was swept by a culture of "convenient," disposable consumer products, particularly those utilizing plastics (many snapshot cameras were already being manufactured from plastics), the stage was set for a disposable 35 mm plastic camera. Fujifilm introduced one in 1986 and Kodak soon followed.

Disposable cameras are popular because they are small, pocketable (thanks to lenses of short focal length), inexpensive, and ultraconvenient. Some models incorporate electronic flashguns, ideal for photographing friends and family socializing indoors. One simply removes the packaging, advances the film to the first exposure, and begins shooting, delivering the entire camera to the processor when the roll is finished. Even owners of expensive cameras purchase disposables if they have forgotten to pack the "good" camera for a trip or want pictures in situations that might place

expensive cameras at risk of damage or loss. The disposable camera uniquely fulfills the almost-forgotten premise of the 1888 Kodak, and suggests that many people found film handling a reason to avoid taking photographs, even with automatic take-up and threading. Displays of disposable cameras near supermarket cash registers attest to their popularity as impulse purchases. Makers emphasized the ''fun'' of using disposable cameras, and users indeed indicated fancy-free feelings. According to the Photo Marketing Association, one-time use cameras are targeted primarily to people aged under 25, and are often used by teens.

Some executives were worried that a low-end new product might have harmed established brands. When Kodak planned its ''single-use'' camera, staff of the company's film division vigorously opposed it, fearing that photographs using inexpensive plastic lenses would compare unfavorably with those from 35 mm cameras. However, it didn't matter. The Kodak Funsaver of 1994 was purchased for specific jobs: people wanted a camera on vacation but either didn't own one or had forgotten to bring one. The Funsaver competed, not with 35 mm cameras, but with nonconsumption, and customers were pleased with the quality.

Disposable cameras have appeared in many designs and styles. In 1989 the single-use Kodak Stretch 35 Camera produced 3.5- by 10-inch (90 by 250 millimeter) panoramic prints, a format previously available only with expensive specialized cameras. The single-use Kodak Weekend 35 camera was an all-weather camera that could take pictures underwater down to a depth of 2.5 meters, and offered low-risk beach photography. Disposable cameras with built-in electronic flash are popular for parties, weddings, and other social events. (The flash unit contains a large capacitor—120 to 160 microfarads, rated for 330 volts.)

Early disposable cameras had poor lenses and films were less advanced, but later disposables were greatly improved. They used molded aspherical lenses, which would be expensive in a traditional glass lens. These shapes enable a simple lens to render sharp images without increased cost, since the camera is entirely molded out of plastic anyway. Disposable cameras do not have a rewind mechanism, so they operate backward compared to a conventional 35 mm camera: as the film is exposed, it is rolled into its canister. In 1998 Polaroid introduced a single-use camera which the user returned to Polaroid for recycling, and the Polaroid JoyCam appeared in 1999. Kodak added a switchable single-use camera with a choice of two print formats and faster flash recharge time in 1998.

Manufacturers, who claim to be committed to recycling, prefer the term ''single-use'' over the popular ''disposable.'' Kodak was especially stung by environmentally minded consumer criticism of its disposable cameras. In response, the company began a ''take-back'' program in 1990 to reuse and recycle the cameras. Kodak's initiative is more extensive than that of other brands. About 86 percent of a camera's weight is recycled or reused. Most of the remaining weight of flash models is the battery, which is reused or donated to charity. The outer covers of Funsaver cameras are recycled. The chassis, basic mechanisms, and electronic flash systems are tested and reused. Components that fail inspection are ground up and added to the raw material for molding new cameras. Used lenses are ground up and sold to outside companies for other products. Once camera parts have been used ten times, they are recycled into new components. Kodak pays photofinishers for used cameras, providing a financial incentive to collect them, although it may be inadequate. Kodak eventually began to reimburse labs for the costs of sorting, storing, and shipping, but processors faced a dizzying array of types, brands, and procedures. At the end of the twentieth century, studies showed that less than half of disposable cameras were actually recycled.

Disposable cameras are excessively packaged, although the paperboard and foil containers can be recycled. In mid-1995 Kodak announced that it had recycled or reused 50 million single-use cameras; and in 1996, 70 million. In 1999 Kodak said it had saved 18 million kilograms of waste by preventing the disposal of 250 million single-use cameras.

In the twenty-first century it appears that the disposable camera, often hailed as the savior of film-based photography against the digital onslaught, may collide with the new medium. Ritz Camera, based in Beltsville, Maryland, was the first photographic chain to nationally launch a fully digital disposable camera, selling for about $11. Ritz claimed the throwaway camera had many features of more expensive digital cameras, including automatic flash and timer and the ability to delete unwanted photos. This product combined two of the fastest-growing segments in photography. According to the Photo Marketing Association, single-use cameras grew more than 15 percent annually over a five-year period, representing about 19 percent of film rolls processed. Digital camera sales grew 23 percent in

2003 and represented almost a third of consumer photography. Ritz's Dakota Digital Camera produced more expensive but inferior images when compared to disposable film camera results. Some predict that professional cameras will ensure the survival of film due to unbeatable versatility and image quality, but disposable cameras will also survive because they are so cheap. In 2002, 170 million disposable cameras were sold in the U.S. Digital camera sales were predicted to eclipse traditional film cameras in 2003, according to the Photo Marketing Association, but disposable film cameras still reigned supreme, with 214 million units sold in 2003.

See also **Cameras, 35 mm; Cameras, Automatic; Cameras, Digital; Cameras, Polaroid**

DAVID E. HABERSTICH

Further Reading

At your disposal. *Photo Technique*, London, April 1998. Five professional photographers testing single use cameras.
Blaker, A.A. To toss or not to toss? Should you toss out disposable 35 mm cameras after a single use? Not necessarily. *Darkroom Photography*, 1990, 48–64.
Coe, B. *Cameras: From Daguerreotypes to Instant Pictures*. Crown Publishers, New York, 1978.
Goldsmith, A.A. *The Camera and Its Images*. Newsweek Books, New York, 1979.
The Columbia Electronic Encyclopedia, 6th edn. Columbia University Press, Columbia, 2003.

Useful Websites

Christie, R.D. The Disposable Camera: http://www.ee.washington.edu/conselec/CE/kuhn/labs/camera/fl.htm
Eastman Kodak Co. History of Kodak: From Glass Plates to Digital Images: http://www.kodak.com/US/en/corp/aboutKodak/kodakHistory/kodakHistory.shtml
Grepstad, J. Disposable Camera Photography: http://home.online.no/~gjon/stretch.htm
History of Photography and the Camera: http://inventors.about.com/library/inventors/blphotography.htm
Modern Photography: http://inventors.about.com/gi/dynamic/offsite.htm?site=http://www.infoplease.com/ce6/ent/A0838872.html
Wells, H. Reloading and Adapting Single-Use (Disposable) Cameras: http://mywebpages.comcast.net/hmpi/Pinhole/Articles/Disposable/Disposable.htm

Cameras, Lens Designs: Wide Angle and Zoom

The function of camera lenses is to refract light rays and bend them to form an image inside the camera of a subject outside the camera, to be captured on a photosensitive surface such as film, or more recently, on photosensitive surfaces in digital cameras. Lens design strategies have always been dedicated to improving the quality and fidelity of the camera's images because basic lenses generally are afflicted by various limitations and imperfections, called aberrations.

The story of lens design has been largely a tale of the struggle to maximize lens performance—seeking to increase the effective aperture size (or speed) and improve resolving power and sharpness—while eliminating the aberrations or imperfections which degrade the images they produce. Probably the first camera lens to be used in nineteenth century photography was the singlet landscape lens suggested by William Hyde Wollaston for camera obscuras in 1812. The lens, produced in a meniscus shape, had some of its faults corrected by experimentally locating the aperture at the optimum distance from the lens, and it was used on simple cameras for many decades. The first design improvement for this basic photographic type was the Chevalier lens for daguerreotype cameras, consisting of two glass elements cemented together.

The first lens of high relative aperture that facilitated portraiture with light-sensitive materials of extremely low sensitivity, such as the daguerreotype, was designed by Josef Petzval in 1840. It contained a telescope objective at the front, widely spaced from another telescope objective at the rear. This basic design was a classic type, utilized widely throughout the history of photography for still cameras, motion picture projectors, as well as microscopes. Originally used for portraiture, the Petzval design was eventually supplanted by other designs.

The most popular camera lens of the latter nineteenth century until 1910 was the symmetrical "duplet" consisting of two identical lens groups on either side of a central diaphragm, because this design eliminated many types of lens aberrations. Other activity in nineteenth century lens design included the use of new types of glass and experimentation with both symmetrical and asymmetrical systems. By 1893 a new basic design, the "triplet," was introduced, consisting of three air-spaced singlets or three air-spaced lens groups. In the twentieth century enormous quantities of triplet lenses were manufactured for moderately priced folding cameras and other cameras. Considerable research went into methods of modifying the triplet design to produce faster, higher-aperture lenses.

The relationship between the focal length of a camera lens and image size has been somewhat arbitrarily codifed over the years. A "standard"

lens is usually considered to be one whose focal length approximates the diagonal dimension of the rectangular image produced by the camera, on the theory that this produces "normal" perspective. This is a simplistic notion, as the perception of perspective is more complex than many realize, but the standard remains. Thus a 35 mm camera is routinely equipped with a lens of about 45 to 55 mm. A lens of shorter focal length is considered "wide angle," although it became fashionable at the end of the twentieth century for photographers who prided themselves on selecting a wide-angle lens as their personal "standard" workhorse. A specially constructed lens of extreme wide angle, called a fisheye lens, has scientific applications, but it was popularized as an experimental tool for seemingly distorted and intentionally bizarre effects. A lens of greater than standard focal length is called "long-focus" or "telephoto." A telephoto lens denotes a particular design type, however, which is intended to produce a lens significantly shorter than its effective focal length would suggest; that is, the total length from the front lens to the image plane is actually less than the focal length. The second principal plane lies outside the positive end of the lens system, or the rear principal plane will be in front of the front component. Not every long-focus lens is a true telephoto, despite the confusing tendency of photographers to use such nomenclature uncritically. Telephoto construction requires a positive front element that is widely separated from a negative rear component. As it is more difficult to correct aberrations in a telephoto lens, a long-focus lens of conventional construction may be preferable; a telephoto is used when the need for compactness equals or outweighs the concern for optimum quality.

Perhaps the single most important contribution of the twentieth century to photographic lenses, especially since it came to be utilized almost universally, was the application of hard, permanent coatings on lenses to reduce the surface reflectivity of the glass. The fact that coatings could reduce reflections was discovered by H. Dennis Taylor in 1896 when he worked with tarnished lenses, but attempts to reproduce such tarnish with precision were unsuccessful. In 1936 John Strong suggested depositing a thin layer of a low-index material such as calcium fluoride onto lens surfaces,. The thin film reduces surface reflections through interference, in which beams of light passing through the coated lens interfere with each other and eliminate the reflected light. Commercial lens coatings were offered in December 1938, and the method was subsequently greatly improved. In addition to the very real technological advance that lens coating represented, the esthetic appeal of coated lenses evidently had an impact on the market. Advertising images of cameras with gleaming lens coatings of subtle, attractive colors—browns, blues, magentas, and purples—helped to make amateur photography one of the most important hobbies of the twentieth century. Lens coatings tended to connote precision and drew upon a ready market of gadget collectors.

A "zoom" lens has some means for continuously varying the focal length of the lens while retaining image focus at the film plane. Originally popular for motion-picture photography for the obvious advantage of both continuous and seamless zoom effects, as well as the ability to change the apparent camera-to-subject distance rapidly without the need to change lenses. They were extremely difficult to design and manufacture with adequate corrections, however, and they did not become plentiful until the mid-twentieth century. The best known zoom lens at that time, the Zoomar, had 22 separate lens elements in the 16 mm motion-picture model and an even larger number in its 35 mm counterpart. In the late twentieth century low-cost manufacture of zoom lenses made them more available to the amateur market, and many 35 mm cameras were routinely sold with variable-focus lenses. Eventually small "point-and-shoot" 35 mm cameras became ubiquitous, and most were equipped with zoom lenses in lieu of interchangeable lens capability.

All modern lenses are corrected achromatically, which means that two different colors will focus at the same distance between the lens and focal plane, but the difficulty increases with faster telephoto and zoom lenses. Apochromatic (APO) lenses are corrected to focus three different colors in the same plane. Such designs were expensive until the 1990s. Engineers rely upon computers to develop improved optical and glass manufacturing technology to provide high-quality low-dispersion glass. The availability of computerized lens design undoubtedly constitutes one of the most important advances in the history of twentieth century photographic optics.

Unusual lenses have been manufactured for special purposes. One of the most striking is the anamorphic lens developed for widescreen CinemaScope motion pictures in the 1950s. This lens could be used to compress laterally a wide-screen image into the normal dimensions (1:1.85 aspect ratio) of a conventional 35 mm movie frame, then expand the image to fill the CinemaScope screen, with its extreme aspect ratio (1:2.35).

Another exciting late twentieth century innovation relating to camera lenses actually has only an indirect relationship to optics. This was the advent of automatic focus or autofocus camera systems: properly speaking, these methods involve the use of a motor to focus the lens for the user, operating via signals received through such technologies as SONAR, infrared pulses, and computer analysis.

See also **Cameras, 35 mm; Cameras, Automatic; Optical Materials**

DAVID E. HABERSTICH

Further Reading

Brandt, H.-M. *The Photographic Lens*. Focal Press, London, 1956.

Kingslake, R. *Lenses in Photography: The Practical Guide to Optics for Photographers*. Garden City, New York, 1951.

Kingslake, R. *A History of the Photographic Lens*. Academic Press, Boston, 1989.

Lockett, A. *Camera Lenses: A Handbook to Lenses and Accessories for Amateur and Professional Photographers*, 5th edn, completely revised by Lee, H.W. Sir Isaac Pitman & Sons, London, 1962.

Neblette, C.B. *Imaging Processes and Materials*, Sturge, J., Walworth, V. and Shepp, A., Eds, 8th edn. Van Nostrand Reinhold, New York, 1989.

Neblette, C.B. *Photographic Lens Manual and Directory*. Morgan & Morgan, New York, 1959.

Neblette, C.B. and Murray, A.E. *Photographic Lenses*, revised edn. Morgan & Morgan, Dobbs Ferry, 1973.

Smith, W.J. *Modern Lens Design*. McGraw-Hill, New York, 1992.

Wildi, E. *Photographic Lenses: Photographer's Guide to Characteristics, Quality, Use and Design*. Amherst Media, Buffalo, CO; and Turnaround, London, 2002.

Useful Websites

Shepperd, R. Marvels of Lens Design, in *Outdoor Photographer*, Nov. 1998. http://www.outdoorphotographer.com/content/pastissues/1998/no/lens.html

Cameras, Polaroid

Edwin Herbert Land and Polaroid, the company that he founded, produced the first self-developing cameras by employing synthesized sheet material that could align light waves. Discovered in 1934, this plastic was dubbed Polaroid for its resemblance (*-oid*) to polarization. Initially, the company used the plastic in sunglasses and filters. Concluding that the sheeting had potential as part of an instant camera, Land assigned a team to develop such a product.

An immediate sensation, the first Polaroid Land camera went on sale in 1948 in an upscale department store. Marketed as a luxury item for amateur photographers, the camera cost more than other types yet quickly found buyers among people drawn by the appeal of instant results. Since instant photography skips the step of submitting film to developers, the Model 95 also found a market among those who wanted to keep intimate photos of lovers away from the eyes of others. Professional photographers often purchased the camera to test setup conditions. More than half a million Model 95 cameras sold in the first five years of production. The camera contained a folding bellows that connected the lens housing to the body, which was covered in imitation leather. A latch popped the camera open and, when locked into the closed position, the sensitive interior was fully protected meaning that no lens cap was necessary. The camera produced 3.25- by 4.25-inch (90 by 110 mm) sepia-toned photographs, eight to a roll. It weighed just over 2 kilograms when loaded with a film roll, making it difficult for many people to handle. An exposure value (EV) number from one to ten described settings for aperture and shutter speed. Polaroid recruited General Electric to design an inexpensive exposure meter to read the light of the scene expressed in the same EV numbers. With the meter clipped to the camera, the photographer checked the meter then set the same number by turning a wheel on the shutter board before taking a properly exposed picture. Within a few years, this EV system would be adopted, with some modifications, as the standard of the amateur photographic industry.

To produce an instant photograph, the camera exposed an image on a roll of negative photographic paper pulled down from the top of the camera into view of the lens. This paper met a set of rollers with positive paper pulled up from the bottom of the camera. Interspersed on the positive roll were small pods of chemicals: the standard developer hydroquinone and the typical fixer sodium thiosulphate. As the papers pulled through the rollers together and out of the camera body, the pressure of the rollers burst the reagent pod, spreading chemicals evenly through the middle of the positive–negative sandwich. The chemical reaction took about sixty seconds to complete, at which time the photographer peeled the positive print from the negative. The process was not foolproof and the film did not always peel easily.

The Model 95 failed to produce consistently good photographs because Polaroid had been

unable to produce a high-quality, easy-to-use film. The company solved this potential customer satisfaction problem by encouraging camera users to think of photography as a creative process filled with trial and error. In 1950, Polaroid introduced black and white film but consumers soon reported a fading problem. After determining that the problem was caused by contaminants present in the air, Polaroid reconstructed the positive print and instructed consumers to add the annoying step of painting pictures with a protective coating. Famed photographer Ansel Adams, serving as a paid consultant, suggested ways to improve the tonal value of Polaroid film and persuaded the company to market 4- by 5-inch (100 by 130 mm) sheet film. The smooth surface of the film showed smudges, picked up glare, and revealed fine-detail flaws in the photograph. Color film, Polacolor, became available in 1963. Polacolor records the three primary colors in three light-sensitive layers. When the film is exposed, some dye is trapped while the remainder transfers to the receiving layer thereby reproducing the color of the subject. The film did not hold its color over the years. Wastage and quality of film remained concerns, and Polaroid was never able to match the quality and consistency of Kodak's equivalent 35 mm film.

While Polaroid worked to perfect its film, it continued to introduce instant cameras. Like other photography companies, Polaroid made most of its profit on film, giving it an incentive to reduce camera prices as much as possible. In 1954, the moderately priced Highlander reflected the company's plan of offering increasingly lower-priced models. The 1963 Automatic 100 had pack film that developed outside the body, so the photographer could take a series of shots without waiting for the film to develop. In 1965, the Swinger became Polaroid's first low-cost model. Named after a slang word for "fun person" to appeal to the teenage market, this camera had a high-impact plastic boxy body. It used roll film that produced small black and white photographs and tended to jam if the camera suffered rough treatment. The Swinger contained an innovative exposure control device. When the user gripped a small red stick that projected above the shutter, an electric bulb inside the camera illuminated a checkerboard display in the viewfinder just above the image of the scene to be photographed. By rotating the control stick, the photographer could open or close the aperture to admit enough light to balance the brightness of the bulb which was keyed to light sensitivity of the film. When the two brightnesses were balanced, the checkerboard spelled YES. Later versions of the Swinger included a T-bar strap to make the camera easier to hold while extracting film.

In 1968, the Big Swinger replaced the Swinger before the Colorpack II succeeded it in 1969. The Colorpack accepted black and white or color film, had an electronic shutter for automatic exposure control, an electric eye, a more precise lens, and a built-in flashgun that relied upon four-shot flashcubes. In 1971 the company added a self-timer that gave the photographer three seconds to enter the picture frame before the shutter snapped. Additionally, a beeper sounded when the picture was ready to be peeled from the negative. A third device also activated when insufficient natural light existed for a good picture. In 1972, SX-70, the first pocket-sized Polaroid camera, used cast plastic technology to fold light through internal lenses and mirrors. The 116 mm lens was slightly wide-angle to make the best compromise among all the possible situations under which camera users might be operating. Portraiture remained difficult since the lens tended to flatten and distort anything filling up the frame at a short distance. To remedy this problem, Polaroid eventually developed a telephoto lens attachment. Later Polaroid cameras featured automatic focus, a built-in flash with an automatic recharge feature, a frame indicator displaying the numbers of pictures remaining in a film pack, and a close-up adaptor.

In the 1980s, the popularity of Polaroid cameras dipped as cheap 35 mm cameras and one-hour photo shops permitted consumers to produce better-quality photos without sacrificing a great deal of developing time. In the 1990s, the advent of digital photography further eroded the market for Polaroids as photographers began to use computers to transmit images although the cameras remained in heavy use for identification purposes, particularly for licenses issued by motor vehicle bureaus.

See also **Cameras, 35 mm; Cameras, Automatic**

CARYN E. NEUMANN

Further Reading

Adams, A and Baker, R. *Polaroid Land Photography*. New York Graphic Society, Boston, 1978.

Carr, K.T. *Polaroid Transfers: A Complete Visual Guide to Creating Image and Emulsion Transfers*. Amphoto Books, New York, 1997.

Langford, M. *Instant Picture Camera Handbook: All About Your Instant Picture Camera and How to Use It*. Alfred A. Knopf, New York, 1980.

McElheny, V.K. *Insisting on the Impossible: The Life of Edwin Land*. Perseus Books, Reading, 1998.

Olshaker, M. *The Polaroid Story: Edwin Land and the Polaroid Experience*. Stein & Day, New York, 1978.

Wensberg, P.C. *Land's Polaroid: A Company and the Man Who Invented It*. Houghton Mifflin, Boston, 1987.

Cameras, Single Lens Reflex (SLR)

The principle of the single-lens reflex (SLR) camera, reflecting the light path by means of an angled mirror behind the lens so as to project an image onto a horizontal glass screen, dates back to the prephotographic portable camera obscura. However, the early photographic experimenters using this type of instrument soon found that the light loss inherent in the arrangement was unacceptable and adopted cameras with a direct light path.

Although the 1839 camera made by Giroux for Daguerre (the earliest photographic camera to be widely sold to the public) had a reflex viewing mirror, the first true SLR camera was patented by the Englishman, Thomas Sutton in 1861. Sutton's camera incorporated an internal mirror that reflected the image formed by the lens up into a horizontal glass screen. The mirror was swivelled up to cover the screen during exposure. The relative insensitivity of the wet collodion plates in use at the time and the delay imposed by the need to prepare them directly before exposure negated the great advantage of the reflex camera, continuous inspection allowing adjustment of focus. Few of Sutton's cameras were made and the reflex principle was almost completely abandoned until "fast" gelatin dry plates were introduced in the 1870s.

The first SLR camera to be at all widely sold was patented by an American, Calvin Rae Smith, in 1884. Called the Monocular Duplex camera, it was fitted with an internal mirror attached to a wedge-shaped unit pierced with an aperture. Raising the unit allowed light from the lens through to a plate at the back of the camera, thus acting as a shutter. The English company, Perken, Son and Rayment, marketed a similar camera in 1888. S. D. McKellen's Detective camera, also sold in the U.K. during the same year, was equipped with a roller-blind shutter fitted between lens and mirror. Encapsulating the advantages of the SLR design, a contemporary review noted that an "exact and full-size picture" could be seen "up to the very moment of firing."

During the last decade of the nineteenth century, several new SLR camera designs were produced. Innovations included cameras designed to use rollfilm or magazines of cut film instead of plates, and models fitted with focal plane shutters. The ability to frame and focus moving subjects up to the moment of exposure led to several models fitted with long focus lenses being advertised as particularly suitable for naturalists or for photographing wildlife. The most influential design of the period was the American Folmer and Schwing Graflex camera of 1898. By the standards of the day this was a compact model, the lens panel being fitted to a bellows extension moved by a rack and pinion mechanism. The English Soho Reflex camera introduced in 1905 and built to a similar pattern was immensely popular and, along with derivatives of the Graflex, was marketed until after World War II. Both types were much favored by press photographers. The amateur market was first widely exploited in the 1920s and 1930s as many relatively inexpensive rollfilm reflex cameras began to appear.

A great drawback of the early reflex camera was its bulk. Folding or collapsible models were produced from as early as 1903. Many of the best examples came from German manufacturers, perhaps the finest being the Miroflex models first produced in the late 1920s. There was a broad trend at this time towards compact all-metal cameras and a notable feature of the period was the success of precision-constructed 35 mm cameras such as the Leica and the Contax. A 1933 reflex-mirror box attachment for the Leica camera has been described as the first application of the reflex principle to 35 mm photography. The first production 35 mm SLR camera was probably the Russian Sport camera of 1935, followed a year later by the much more influential and widely sold Kine Exakta by Ihagee of Dresden. Although basically a 35 mm version of a 127 rollfilm camera, the Kine Exakta was a well-made instrument equipped with a fast standard lens and built-in flash synchronization.

The development of 35 mm SLR cameras to the point where they displaced rangefinder models derived from the Leica and Contax as the serious photographers choice was primarily due to the introduction of the pentaprism viewfinder. For the first time, it made possible a compact reflex camera that allowed right-way-up, right-way-round, eye-level viewing. More than one pentaprism patent was filed in the 1940s but the most influential production, SLR camera with a pentaprism viewfinder was the Contax S of 1948. The Contax was soon followed by a series of 35 mm pentaprism reflex cameras produced increasingly in Japan.

From the1950s, the evolution of the 35 mm pentaprism SLR camera continued by way of a series of innovations made possible by advances in light sensor design and electronic components and circuitry. Many important features were pioneered

in European cameras but were developed and refined by Japanese companies. Asahi Pentax models of the 1950s were the first widely sold SLR cameras to incorporate modern instant return mirrors and automatic diaphragms. Similar features appeared in the modular-constructed Nikon F camera of 1959. When later equipped with through-lens metering and marketed with a wide range of lenses and accessories, the early Nikon F was arguably the most influential camera of the period. The first SLR cameras with through-lens metering (the microelectronic coupling of exposure meter to shutter and diaphragm to provide fully automatic exposure control) offered for sale to the public were again Japanese products: the Topcon RE Super of 1963, closely followed by the Pentax Spotmatic. Since 1970, the quality camera market has been dominated by the versatile, Japanese manufactured, fully automatic, 35 mm single-lens reflex model, such as the Canon A and Nikon F series cameras. Perhaps the only exception is the square picture format camera with interchangeable film magazine exemplified by the 500C Hasselblad, favored in certain branches of professional photography.

See also **Cameras, 35 mm; Cameras, Automatic; Cameras, Lens Designs, Wide Angle, Zoom**

JOHN WARD

Further Reading

Coe, B. *Cameras: From Daguerreotypes to Instant Pictures*. Marshall Cavendish, London, 1978.

Coe, B. *Kodak Cameras: The First Hundred Years*. Hove Foto Books, Hove, 1988.

Condax, P.L, Tano, M., Takashi, H. Fujimura, W.S. *The Evolution of the Japanese Camera*. International Museum of Photography, Rochester, 1984.

Gaunt, L. *Canon A Series*. Focal Press, London, 1983.

Gernsheim, H. and Gernsheim, A. *The History of Photography*. Thames & Hudson, London, 1969.

Hicks, R. *A History of the 35 mm Still Camera*. Focal Press, London, 1984.

Janda, J. *Camera Obscuras. Photographic Cameras 1840–1940*. National Museum of Technology, Prague, 1982.

Kingslake, R. Folmer and Schwing-Graflex, in *The Photographic Manufacturing Companies of Rochester*. New York. Rochester, 1997.

Lothrop, E.S. *A Century of Cameras*. New York, 1973.

Reflex Camera, in *The Focal Encyclopedia of Photography*, vol. 2, fully revised edn. Focal Press Ltd, London, 1965.

Cancer, Chemotherapy

When German bacteriologist Paul Ehrlich coined the term "chemotherapy" early in the twentieth century, he referred to drugs that had affinities for certain cells, much as nature's "magic bullets"—antibodies—targeted pathogens. The notion of bullets and targets remained a popular image in anticancer chemotherapy for about 50 years, and it fit especially well into the so-called "war on cancer" during the 1970s and 1980s in the U.S. Subsequently, however, researchers found that anticancer drugs did not kill tumor cells directly but rather induced them to die naturally in a process called apoptosis, or programmed cell death. This later concept, based mainly on nucleic acid studies of cell biology, shifted the paradigm of anticancer therapy from models of war to those of communication, and resulted in 2001 in the first signal transduction inhibitor (STI) drug with many more drugs of that type under development. The STIs act by inhibiting the transmission of a signal, in the form of messenger ribonucleic acid (RNA), from the chromosomal gene to where peptides (preproteins) are assembled, and thus blocking the gene's expression in a cancer mode.

As the twenty-first century began, about 50 anticancer drugs (grouped by the mode of action as alkylating agents, antimetabolites, antibiotics, enzymes, plant alkaloids, and biologics and usually prescribed in various combinations) were effective as primary or adjunct treatments in patients with localized or metastized (tumors spread to a distant site) disease.

Modern chemotherapy, as it developed during and after World War II, was made possible by trends in science, medicine, and economics. They included

1. An accumulation of knowledge in science (pharmaceuticals, tissue culture, inbred mice for tumor lines, microscopy, and basic science)
2. Available capital, largely from the pharmaceutical industry's financial success with antibiotics
3. An apparent need as seen in the relative decline of heart disease mortality and the rise of cancer deaths
4. The social acceptability of pursuing "cures" over "prevention"

The first drug to arise in this matrix was the alkylating agent mechlorethamine (trade name Mustargen) in 1947. Its development resulted from the studies of Alfred Gilman and colleagues at Yale University carried out in 1942 on the effects of nitrogen mustard gas on lymphoid tissue, including lymphosarcomas, or malignant tumors. As a group, alkylating agents work within cells by attaching to and breaking DNA (deoxyribonucleic

acid) strands, interfering with the cell's replication and protein synthesis, and thereby inhibiting cancer cell growth. This action was further demonstrated in such drugs as busulfan (Myleran) in 1953 for chronic myeloid leukemia; chlorambucil (Leukeran) in 1953 for chronic lymphocytic leukemia; melphalan (Alkeran) in 1953 for multiple myeloma; and cyclophosphamide (Cytoxan) in 1957 for lymphomas.

Generally, anticancer drugs worked best against fast-growing "liquid" cancers such as the leukemias and lymphomas and were less effective against slow-growing "solid" tumors that metastasized early. While the overall cancer cure rate using various chemotherapies was around 15 percent, some drugs showed spectacular results. Mercaptopurine (Purinethol), the first antimetabolite chemotherapy in 1953, virtually reversed the mortality from 80 percent to 20 percent for children suffering from acute lymphocytic leukemia. Antimetabolites work at the cellular level by substituting DNA base analogs that interfere with the cell's chromosome synthesis and prevent it from replicating. Gertrude Elion and George Hitchings, who pioneered this class of drugs in the 1940s, shared the 1988 Nobel Prize in chemistry. By the time they received the prize, antimetabolite drugs were also indicated as a treatment for HIV (human immunodeficiency virus) infection.

Antibiotics constituted a third type of cancer chemotherapy, beginning with Harvard University researcher Sidney Farber's 1954 clinical studies of dactinomycin (Actinomycin-D) among children with Wilm's tumor (kidney cancer). As investigators around the world discovered more species of streptomyces, they learned that antitumor antibiotics had less toxic side effects than antibacterial antibiotics and mainly worked by inhibiting DNA and RNA (ribonucleic acid) functions and protein synthesis. The array in this class grew and included bleomycin (Japan, 1965), daunorubicin (Belgium, 1964), doxorubicin (South Africa, 1969), mitomycin (Japan, 1956), and mitoxantrone (Germany, 1982). Similarly, investigators in the early 1950s learned that enzymes, particularly asparaginase, restricted tumor-cell proliferation by blocking protein synthesis. It was marketed under the trade name L-Asparaginase in 1953.

Anticancer plant alkaloids (plants containing nitrogen and usually oxygen, especially in seed plants), first discovered during the 1960s, are actually among the oldest known medicines. The *Leech Book of Bald* from tenth century England mentioned the mandrake plant, also known as May apple, as a cathartic. It also contains a toxin that was discovered in 1970 to act as a mitotic inhibitor and preventor of cell division in resistant testicular tumor cells and small cell lung cancer. By then, however, the *Vinca* plant derivatives vincristine (Oncovin, 1961) and vinblastine (Velban, 1965) had already proved able to bind microtubes during metaphase cell division, thus restricting mitosis. Curiously, another alkaloid (paclitaxel) isolated from the bark of the Pacific yew tree was shown to have a similar effect by a completely opposite action. In 1979 Susan Horowitz of the Albert Einstein College of Medicine in New York demonstrated that paclitaxel promoted so much polymerization during metaphase microtube elongation that it paralyzed phase transition, thereby stopping growth. In 1994, after the politically contentious Pacific Yew Act of 1991 protected the natural source *Taxus brevifolia*, the drug was manufactured in a semisynthetic process under the trade name Taxol and indicated for use in treating ovarian and breast cancer. Hundreds of plants with anticancer properties were subsequently discovered—as the Chinese Institute of Medicinal Plants most recently catalogued—but most exceed acceptable toxicity for most patients.

Among the increasing number of biologic anticancer agents (i.e., those produced naturally by mammalian cells rather than synthesized) are: antiestrogenic hormones such as tamoxifen; interferon, which works by stimulating natural killer cells and macrophages against tumors; interleukin-2, a T-lymphocyte growth promoter; the cytotoxic protein called tumor necrosis factor; and monoclonal antibodies, which can detect tumors as well as deliver small molecule drugs or radioactive isotopes by targeting human tumor antigens. Also in this group of biologic agents are the STIs, the first of which was marketed as Gleevec in 2001 in the U.S. (elsewhere named Gilvec) and approved for treating chronic myeloid leukemia. Administered as a pill, Gleevec targeted tyrosine kinase, which the DNA transposition in chronic myeloid leukemia activated, causing a 10- to 25-fold increase in white blood cells. By inhibiting the protein kinase, Gleevec showed dramatic therapeutic results with minimal side effects. This led many researchers to speculate that protein kinase "cocktails" could have a similar role in making cancer a manageable disease with much reduced mortality, similar to that of the protease inhibitors used in treating AIDS. By early 2002, over 170 other signal transduction inhibitors were in various stages of testing.

During the last decade of the twentieth century, the Human Genome Project produced a "rough

draft" map of all human genes. As that project is completed in the first decade of the twenty-first century, "patient discovery" may well drive chemotherapy research more than "drug discovery." As gene expression analysis creates a molecular classification of cancers and identifies both the individuality of the disease and the predictable efficacy to toxicity response in each patient, "new" anticancer drugs may come from novel discoveries, or from finding precise applications for drugs that have been known since medicine began.

See also **Cancer, Radiation Therapy; Hormone Therapy; Medicine**

G. TERRY SHARRER

Further Reading

De Vita Jr., V.T. Principles of chemotherapy, in *Cancer Principles & Practice of Oncology*, Da Vita Jr., V.T., Hellman, S. and Rosenberg, S.A., Eds., 4th edn. J.B. Lipincott, Philadelphia, 1993.

Laszlo, J. *The Cure of Childhood Leukemia: Into the Age of Miracles*. Rutgers University Press, 1995.

Porter, R. *The Greatest Benefit to Mankind: A Medical History of Humanity*. W.W. Norton, New York, 1997.

Pratt, W.B., Ruddon, R.W., Ensminger, W.D. and Maybaum, J. *The Anticancer Drugs*, Oxford University Press, New York, 1994.

Waalen, J. Gleevec's glory days. *Howard Hughes Medical Institute Bulletin*, 14, 5, 10–15, 2001.

Cancer, Radiation Therapy

Prior to the advent of the x-ray and radioactive isotopes, cancers were treated by removing them surgically. Many cancers grew undetected because imaging techniques had not yet been developed. The Roentgen ray, later referred to as the x-ray and named for the German physicist and discoverer Wilhelm Conrad Roentgen in the late nineteenth century, would change both the diagnosis and treatment of cancer. (See X-Rays in Diagnostic Medicine.)

Skin damage arising from careless use of x-rays led to early ideas for the therapeutic use of radiation on human tissue, though its ionizing effect was not understood at the time. Today scientists understand that some cancer cells are more susceptible to damage from ionizing electromagnetic radiation than are ordinary cells. The therapeutic use of x-rays began in 1896 when Emil Grubbe used x-radiation to treat a patient with breast cancer. In 1902, Guido Holtzknecht created a chromo-radiometer, which could record and measure the radiation dose administered to a patient, paving the way for systematic use of radiation therapy. One of the first applications of x-ray therapy was to treat ringworm infection. If enough radiation was applied to kill the fungus, hair on the skin fell out, but if too much was applied, the hair did not grow back and the skin was burned. Thus the nemesis of treatment was defined: the narrow margin between enough and too much. By the 1920s, however, x-ray machines were routinely used in hospitals for clinical treatment.

The French physicist, Henri Becquerel is credited with the discovery of natural radioactivity when he observed that uranium salts produced images on nearby photographic plates. Other radioactive elements, first polonium and then radium, were discovered by French chemists Marie and Pierre Curie. Perhaps the first medical application of radium occurred when Marie's husband Pierre burned his arm with it. The Curies lent radium to Paris physicians, and one early documented case in 1907 described the removal of a child's facial angioma using a crossfire technique.

The first x-ray therapy was administered using radon gas in tubes. The challenge was then and continued to be the manipulation of a ray (which is straight) into a body that has contours, many layers, and organs of different densities and sensitivities. The objective was to obtain adequate tissue depth with rays without destruction of surrounding cells.

In 1913, the first external beam machine used a cathode, or "Coolidge," tube for treating superficial tumors. It was referred to as an orthovoltage applicator and was very slow. Although these machines were extremely limited, they introduced a new genre of technology to medicine known as x-ray therapy, later called radiation therapy. In the 1930s, the use of so-called radium bombs (telecurietherapy) led to refined treatment times, shielding to avoid exposure of healthy tissue, and prescribed doses of radiation. However, results at that time were only palliative.

The Van de Graaff generator, built in 1931, was able to build up a high electrostatic charge and thus high voltages (up to 1 million volts). A medical Van de Graaff, in which electrically accelerated particles were used to bombard atoms and produce radiation, was first used in a clinical setting in 1937. Throughout the 1940s, radiation treatment with the Van de Graaff allowed a very narrow, targeted beam, higher energy (about 2 megavolts), and less treatment time than with gas tubes. But as with most other cancer treatments of the 1940s, these efforts were still palliative and only

temporarily relieved pain or reduced the size of a tumor.

The first circular electron accelerator, named the betatron, was built in 1940 by Donald Kerst and Robert Seber. Originally designed for research in atomic physics in the U.S., the betatron was soon adopted for clinical use. Its first clinical application was by Konrad Gund in 1942 in Germany during World War II, and it was first used by Kerst in the U.S. in 1948. Both directly produced electrons, and x-rays produced by accelerators were an ideal source for therapy and a considerable improvement over the energy that could be achieved with gas and vacuum tubes (higher energy rays have better penetration properties). The betatron energy range of 13–45 megavolts, with 25 megavolts being optimal for therapy, made the device suitable. Linear electron accelerators were developed simultaneously by D.W. Fry in England and William Hansen in the U.S. The first patient to be treated in London with a linear accelerator was in 1953.

Until the new specialty of radiation oncology was recognized in the 1960s in the U.S., diagnostic radiologists administered radiation therapy. Subsequently, the European Society for Therapeutic Radiation and Oncology as well as many other organizations of these specialists have formed worldwide.

In the early 1950s, a group of Canadian scientists isolated a highly radioactive cobalt-60 isotope from a nuclear reactor. This provided a source of gamma rays, popularly and misleadingly referred to as a "cobalt bomb," which could be directed at patients. The Cobalt 60 machine emitted gamma rays of 1.25 megavolts at a distance of 50 to 60 centimeters, and could penetrate deep tissues. Because of the danger of exposure to these rays, buildings in which the machines were located were required to have walls of very thick lead. Many of the original cobalt gamma ray systems have been replaced with linear accelerators.

In 1975, the development of proton beam radiation allowed for higher doses of radiation to target tissues while sparing adjacent cells. Since that time much of the progress in radiation therapy has been through the application of other technologies. Refinements have included more stable machines, radiation at higher rates, modifications to the treatment table, mobility of various machines, higher energy outputs, and collimators (a device to direct the beam). Energy is now described in millions rather than thousands of electron volts. Most machines treat patients in the range of 10 to 25 megavolts or 18 to 20 megavolt photons. The new collimator takes the place of lead positioning blocks that were previously limited in shape and size. One system consists of 25 moving parts that can shape the direction of the treatment to conform to the target tumor.

Miniaturization of technology has allowed for as many as 120 motors to fit in the head of certain machines to deliver radiation to the patient. The newest system of external beam radiation is IMRT (intensity modulated radiation therapy), which links the treatment planning system to the linear accelerator and the multileaf collimator. IMRT has reduced the amount of radiation to surrounding tissues and provided high-resolution images of the patient's anatomy.

By the 1990s, highly sophisticated imaging technology with the use of computed tomography (CT) scans, magnetic resonance imaging (MRI), and ultrasound facilitated more accurate treatment planning for radiation oncology. Radionuclides delivered to bone tissue can identify malignant tissues in a bone scan. More recently, the use of positron emission tomography (PET) has enabled physicians to image metabolic processes and to track tumor metastases from lung and bone tumors.

Another form of radiation therapy, brachytherapy, uses application of a source to tissues a short distance away. Typical sites are the lung, where high doses of radiation are given over a two-week period through a catheter placed in the lung, and the cervix, where cesium-137 is placed in the vagina for a few hours. Radioactive iodine and palladium are used to treat prostate cancer by placing or implanting "seeds" (tiny titanium cylinders containing the radioactive isotope) in the gland, and radioactive palladium is also used in implants for tumors of the tongue. This is known as implant therapy.

One tumor site that posed the most difficult problems for both external beam radiation therapy and implant technology was the brain because it is covered by bone. The Gamma Knife, developed in 1968 by Lars Leskall and Borge Larsson in Sweden, is an instrument that delivers a concentrated radiation dose from Cobalt-60 sources. It fires 201 beams of radiation into the skull that intersect at the target site. No single beam is powerful enough to harm surrounding tissue, but the cumulative effect of this precision tool destroys the tumor.

See also **Cancer, Chemotherapy; Cancer, Surgical Techniques; Nuclear Magnetic Resonance (NMR, MRI); Particle Accelerators: Cyclotrons, Synchrotrons, Collider; Particle Accelerators,**

Linear; Positron Emission Tomography (PET); Tomography in Medicine

LANA THOMPSON

Further Reading

Khan, F. *The Physics of Radiation Therapy*. Williams & Wilkins, Baltimore, 1995.

Mould R.F. *A Century Of X-Rays and Radioactivity in Medicine*. Institute of Physics Publishing, Bristol, 1993.

Oldham, M. Radiation physics and applications in therapeutic medicine. *Physics Education*, 36, 6, 2001.

Rosenow, U. 50 years electron therapy with the betatron. *Zeitschrift für Medizinische Physik*. 9, 2, 73–76, 1999.

Thomas, A., Isherwood, I. and Wells, P.N.T., Eds. *The Invisible Light 100 Years of Medical Radiology*. Blackwell Science, Oxford, 1995.

Cancer, Surgical Techniques

If tuberculosis, with its treatments of rest and fresh air, was the dominant scourge of the nineteenth century; cancer, with its dramatic and often devastating surgical and pharmaceutical interventions, became the dreaded curse of the twentieth. Medical understanding of the body had shifted from a humoral framework in which disease and illness, including tumors, were believed to be caused by imbalances of bodily fluids, to an anatomical–pathological framework in which tumors were understood as solid aberrations to the structure of a particular organ. This shift, coupled with breakthrough developments of anesthesia, antisepsis and asepsis, dramatically increased the use of surgical intervention as a means of treating patients with cancer.

In the early part of the twentieth century, treatment of cancer tended toward radical surgery due to the belief that complete removal of the offending tumor—along with surrounding tissue, and often the entire organ—would prevent a recurrence of the cancer. The results were often hideous disfigurement and poor quality of life—many times with only marginal gains in length of survival.

In treating breast cancer, William S. Halsted (1852–1922), who spent much of his career at the Johns Hopkins Hospital in Baltimore, Maryland, was one of the earliest proponents of the radical mastectomy. Halsted reported his surgical findings to the American Surgical Association in 1898 and 1907. At the latter meeting, Halsted reported that two fifths of his patients who were followed more than three years after their operation were considered to be three-year cures. Unfortunately, two thirds of his patients died of breast cancer despite the radical surgery. Notwithstanding these dismal statistics, Halsted's approach was adopted as standard treatment due in large part to the fact that he trained numerous surgeons at Hopkins who perpetuated his technique of radical mastectomy as the first line of intervention for breast cancer. Even early on, however, the radical mastectomy endured its share of criticism, particularly from the English surgeon Geoffrey Keynes, who decried its disfiguring effects.

Despite criticism of radical surgery, surgical removal of cancers became even more aggressive by midcentury, leading to the "super-radical" mastectomy popularized by such figures as Owen Wangensteen of the University of Minnesota and George T. Pack of the Memorial Sloan-Kettering Hospital in New York City. Similarly, total excisions of lung and stomach cancers became the norm by the 1950s. Gynecological surgery followed a similar path with super-radical surgeries of the uterus and vagina being the standard treatment at the turn of the twentieth century. Medical luminaries such as Drs. Christian Albert Theodor Billroth, Ernst Wertheim, and Alexander Brunschwig were themselves early proponents of this kind of radical surgery, despite statistics of high mortality. By midcentury, a number of radical surgeries became commonplace: hemicorporectomy (surgical removal of the lower half of the body), hemipelvectomy (amputation of a lower limb through the sacroiliac joint), super-radical mastectomy, complete pelvic exenteration (removal of internal organs and tissues), and surgical removal, or resection, of head and neck tumors. The search for a cure for cancer as well as the radical surgical interventions were replete with military metaphors—the war on cancer, a crusade for better health, battling the illness, advances in research—which characterized the aggressive assault on the disease while implicitly justifying the physical carnage that often resulted.

In the last few decades of the twentieth century, however, surgical techniques for the treatment of patients with cancer became more sophisticated and refined and also became more responsive to consumer demands. Ultraradical cancer surgery with its mutilating effects was, to a great extent, discontinued. Where radical procedures were necessary, innovative developments in plastic and reconstructive surgery, coupled with intensive physical and occupational rehabilitation programs, strove to offset and diminish disfigurement and disability as lingering postsurgical effects. Methodologies to evaluate the morbidity and mortality of surgery patients improved during the

twentieth century despite the fact that the gold standard double-blind research approach is not ethically applicable to surgery patients, in contrast to medical patients.

Surgical treatment of cancer is increasingly performed in concert with other kinds of treatment modalities, such as chemotherapy, radiation, hormonal therapies, psychosocial support, and rehabilitation. This sharing of responsibility for the cancer patient among surgeons and other specialists also illustrates the much greater input that nonsurgeons have with the care of cancer patients.

Surgical treatments of cancers have proven to be most effective when the tumor is discrete and located within one segment of the body. Other surgical treatments of cancer include preventive surgery (a partial or total removal of an organ where there is a high risk that a cancer will develop), diagnostic procedures (such as biopsies), cytoreductive surgery (where most of the cancer is removed surgically and the remainder is treated with radiation or chemotherapy), and palliative surgery (where the goal is to alleviate the symptoms of the cancer, rather than curing the cancer itself).

In the waning years of the twentieth century, laser surgery evolved as the latest innovation in treating cancer. The appealing possibilities of laser surgery echo those of the early decades of the century, when x-ray and radium treatments held the promise of removing the cancer without invasive surgery, resulting scars, or the possibility of disturbing the tumor site in such a way as to prompt cells to proliferate and metastasize. However, radiation-induced injuries—including edema, sterility, flesh burns, and even death—and the difficulty in limiting radium treatments specifically to cancer cells without damaging the normal cells, posed challenges to these nonsurgical treatments. Laser surgery seems to answer these challenges by offering highly specific targeting and destruction of cancer cells, minimal invasiveness to the body, and a high degree of control by the surgical team. By the early twenty-first century, leading academic medical centers were running clinical trials to gauge the effectiveness of laser surgical intervention on the treatment of various cancers.

Conventional cancer therapy in the U.S. continues to revolve primarily around the three disciplines of surgery, oncology, and radiation therapy. However, with patient concerns shifting from mere survival of the disease to the quality of life during treatment of the disease, unconventional therapies have emerged as a complement to the standard approaches. Often referred to as alternative, complementary, or unorthodox therapies, they include nutritional therapies, acupuncture, meditation, prayer, therapeutic touch, and herbal supplements to name but a few. Although the debate continues—particularly in the U.S.—over the efficacy of such therapies, patients often seek to integrate some combination of them along with conventional means.

See also **Anesthetics; Cancer, Chemotherapy; Cancer, Radiation Therapy; Medicine; Surgery, Plastic; Surgery, Keyhole and Micro**

KAYHAN PARSI

Further Reading

Cantor, D. Cancer, in *Companion Encyclopedia to the History of Medicine*, Bynum, W.F. and Porter, R., Eds. New York, Routledge, 1993, pp. 537–561.

Lerner, B.H. *The Breast Cancer Wars : Hope, Fear, and the Pursuit of a Cure in Twentieth-Century America*. Oxford University Press, New York, 2001.

Lerner, M. *Choices in Healing: Integrating the Best of Conventional and Complementary Approaches to Cancer*. MIT Press, Cambridge, MA, 1994.

Patterson, J.T. *The Dread Disease: Cancer and Modern American Culture*, Harvard University Press, Cambridge, MA, 1987.

Rettig, R.A. *Cancer Crusade: The Story of the National Cancer Act of 1971*. Princeton University Press, Princeton, 1977.

Rutkow, I.M. *American Surgery: An Illustrated History*, Lippincott-Raven Publishers, Philadelphia, 1998.

Trohler, U. Surgery (modern), in *Companion Encyclopedia to the History of Medicine*, Bynum, W.F. and Porter, R., Eds. Routledge, New York, 1993, pp. 984–1028.

Cardiac Pacemakers and Valves, *see* **Cardiovascular Surgery and Implants: Pacemakers and Valves**

Cardiovascular Disease, Diagnostic Methods, *see* **Angiography; Electrocardiogram; Nuclear Magnetic Resonance (MNR) and Magnetic Resonance Imaging (MRI); Tomography in Medicine; Ultrasonography in Medicine**

Cardiovascular Disease, Pharmaceutical Treatment

By the end of the twentieth century, deaths from heart disease topped the list of causes of death in the U.S. with a near seven-fold increase over the previous 100 years to 268 annually per 100,000 of total population. In 1900 only two effective drugs were available for heart disease. Digitalis (which slows the heart when abnormal rhythms are present) had been discovered by William Withering in 1775 when a lady suffering from dropsy was treated

successfully by folk doctors with herb tea containing foxglove leaves. Nitrates (which dilate the coronary arteries) had been used for angina (cardiac pain) since 1879. Surprisingly, both these drugs remained in use at the end of the century.

By 1946 it had become apparent that hypertension, or high blood pressure, was the prime predisposing factor in the development of arteriosclerosis and therefore myocardial infarcts (heart attacks). Although a class of drugs called the prostaglandins had been introduced in 1930 for blood pressure control, they were largely ineffective. The ganglion blockers such as guanethidine caused intolerable side effects, and it was not until the 1960s that effective therapy that was well tolerated by patients became available.

Beta-adrenergic blockers

Developed in the 1960s by Sir James Black, the so-called beta blockers such as propanolol and atenolol have an inhibitory effect on adrenaline and so slow the heart and reduce both blood pressure and cardiac contractility. These actions reduce myocardial oxygen requirements. They are an effective treatment for prevention of angina, hypertension, some disorders of cardiac rhythm, and heart disease secondary to an overactive thyroid. There is evidence of both primary and secondary prevention of cardiac infarcts (when a coronary artery is blocked off and a portion of heart muscle dies when deprived of oxygen.

Calcium channel blockers or antagonists

Introduced in 1968, these drugs induce relaxation of smooth muscles in blood vessels, including the coronary arteries which then dilate, by impeding the inward movement of calcium ions. By reducing vascular resistance, the blood pressure falls. They also have an effect on some abnormalities of cardiac rhythm. This class of drugs includes nifedipine, verapamil, and diltiazem.

Fibrinolytic drugs

The end-point of coronary artery thickening and blockage is a myocardial infarct. A series of controlled studies in the early 1980s (including the GISSI Trial of 1983 in Italy) showed the value of fibrinolytic therapy. These are drugs that break down the protein fibrin, which is the main constituent of blood clots. Streptokinase was the agent first introduced, and tissue Plasminogen Activator (tPA) is now the preferred agent as it is fibrin-specific. If given intravenously as soon as possible after an acute myocardial infarction, mortality is significantly decreased and damage to the heart muscle minimized. The drug must be given quickly, hence the emphasis in emergency departments on "door-to-needle time." Survival following an infarct is also improved by the use of the appropriate antiarrhythmic drugs early in treatment. Despite these therapeutic advances, in-patient mortality is between 10 and 20 percent.

Angiotensin-Converting Enzyme (ACE) Inhibitors

The first of these drugs was captopril, introduced in 1980 and soon followed by many others, all similar in action. They inhibit the action of the enzyme which converts angiotensin I to angiotensin II. This peptide is a potent vasoconstrictor and also stimulates the release of aldosterone and vasopressin, both of which increase the blood pressure. Initially, ACE inhibitors were introduced for the treatment of hypertension, but it soon became apparent that they were beneficial for patients in heart failure where the renin–angiotensin–aldosterone system has been activated. ACE inhibitors are now a crucial component in the treatment of this condition. The effects are related not only to lowering the blood pressure but also to the peripheral vasodilation, which has an "off-loading" effect in reducing the cardiac workload. They also have an effect on remodeling the musculature of the left ventricle of the heart, which becomes thickened in hypertension and some cardiomyopathies (primary disorders of the cardiac muscle). In diabetics, even with normal blood pressure, these drugs exert a strongly renoprotective action which has been shown to delay the onset of kidney failure. Many physicians believe that the ACE inhibitors are among the most important therapeutic advances of the twentieth century.

Diuretics

The treatment of the accumulation of fluid in the legs, abdomen, and lungs is central to the treatment of the failing heart. There was no effective therapy until 1919 when a mercurial compound used intravenously for the treatment of syphilis was found to induce diuresis (excretion of large quantities of urine). Mercurial diuretics remained the only therapeutic option until effective oral diuretics were developed from 1957 onward. The first of two main groups of drugs were the thiazides, which block the reabsorption of sodium and potassium in the renal tubules. The "loop" diuretics, such as frusemide, also inhibit reabsorp-

tion of salts, but they act in the loop of Henle in the kidney. All the diuretics can cause serious electrolyte imbalances and frequent monitoring of blood chemistry is mandatory.

Lipid-lowering agents

By 1980 the Kerala project in Finland had shown that intervention in a population with a high death rate from coronary artery disease associated with high cholesterol levels was very effective, and the Framingham study in Massachusetts had been delivering the same message (among many others) since 1945. The ion-exchange resin cholestyramine worked, but patients found it difficult to take. More effective and far better tolerated were the statins (e.g., simvastatin). This class of drugs acts to inhibit an enzyme essential for the synthesis of cholesterol.

Antiplatelet drugs and anticoagulants

Aspirin can be described as an old drug reinvented because of the observations, 100 years after its introduction as an analgesic, that it interferes with the activation of platelets, which are essential to the blood clotting process. It reduces mortality if taken at the onset of an acute coronary event and reduces the risks of further infarction if taken long term. Newer antiplatelet drugs include ticlopidine. Anticoagulants such as warfarin or heparin are more controversial and are not as widely used in the acute phase of myocardial infarction as they were in the 1950s and 1960s. Heparin, however, is of great importance in the crescendo angina syndrome and warfarin is used to prevent blood clots breaking off from the heart and moving around the circulation. This can occur in several circumstances when blood flow through the cardiac chambers is reduced.

By 2000, patients who had been spared the ravages of the tuberculosis and pneumonia that killed so many in 1900 had available a cocktail of drugs to control blood pressure, angina, cholesterol, heart failure, and hinder blood clotting. However, they were getting older, fatter, less active, and were often still smoking.

See also **Cardiovascular Disease, Diagnostic Methods; Cardiovascular Surgery**

JOHN J. HAMBLIN

Further Reading

Fleming, P.A. *Short History of Cardiology*. Rodopi BV Editions, 1996. Not that short, very detailed.

Jowett, N. and Thompson, D. *Comprehensive Coronary Care*. Scutari, 1995. Comprehensive & detailed.

Laurence, D.L. *et al. Clinical Pharmacology*. Churchill Livingstone, London, 1997. Standard textbook Good historical perspective.

Opie, L. *Drugs for the Heart*. W.B. Saunders, Philadelphia, 2000. Very detailed account of current practice.

Timmis, A. and Nathan, A. *Essential Cardiology*. Blackwell, London: 1997. Short textbook, covers all fields of current cardiology.

Cardiovascular Surgery, Pacemakers and Heart Valves

The living, pulsing heart has long awed, yet intimidated, physicians and surgeons. Until recently, a disrupted heartbeat or an occluded heart valve brought certain death for patients, while various kinds of congenital heart malformations killed thousands of children every year.

At the turn of the century, the development of techniques for suturing heart wounds created initial interest in the surgical treatment of mitral stenosis, a narrowing of the mitral valve opening. In 1902, the English physician, Sir Lauder Brunton observed:

> "One is impressed by the hopelessness of ever finding a remedy [for mitral stenosis] which will enable the auricle to drive the blood in a sufficient stream through the small mitral orifice, and the wish that one could divide the constriction as easily during life as one can after death. The risk that such an operation would entail naturally makes one shrink from it."

Braunton's own attempt at a surgical cure (proposed in 1902, but not attempted until 1923) was unsuccessful as were most other early efforts. In 1925, however, surgeon Henry Souttar of the London Hospital, opened a mitral valve using only his gloved finger to separate the valve leaflets. The 19-year-old patient recovered. Souttar's achievement was both rare and precious. By 1929, published reports of 12 patients surgically treated for mitral stenosis showed only two survivors.

Likewise, physiologists in the late nineteenth century understood the electrical nature of the heart's contractions and developed rudimentary measuring devices using electrodes to create crude electrocardiograms that measured electrical activity in heart muscle. The new technology, however, required intricate interpretation before it could be successfully employed as a serious diagnostic tool. By 1907, physiologists had a basic understanding of what caused the heart to beat—the electrical origin in a bundle of tissue in the upper right atrium (the sinus node) and its conduction through

the ventricle. When this signaling went awry, clinical symptoms became apparent. Irregular heartbeats constituted one set of problems. Another was evident in patients who had either a partial or complete "block" in conductivity between the upper and lower chambers of their hearts. Such patients had a clinical profile of a slow pulse accompanied by what were originally described as "fainting fits." These "Stokes–Adams" attacks indicated that the ventricle was not providing an adequate supply of oxygenated blood to the brain. It was this condition, moreover, that inspired researchers to develop electric stimulators for the heart.

Before World War II, however, improvements in blood transfusion, anesthesia, and suturing techniques for blood vessels had begun to make surgery safer and led to the expansion of efforts to operate on the living heart. Combat itself finally dispelled the concept of the heart as a fragile organ, however, when it became possible to remove shrapnel and close the wound to a beating heart without killing the patient. The stimulus for an artificial heart-pacing device as well as advances in valve surgery came after World War II and the development of open heart surgery.

In 1947, Claude Beck of Western Reserve University successfully used an electrical defibrillator internally on a patient whose heart had stopped during routine surgery. In 1952, Paul M. Zoll announced that he had kept a Stokes–Adams patient alive through a bedside defibrillation device. Zoll's most important insight was recognition that he should pace stimulate the patient's ventricle, rather than the atrium, but his treatment, employing 130 to 150 volts was too painful for extended patient use. Implanting part of the pacemaker would lower the voltage requirement, and this idea was pursued at the University of Minnesota Medical School. In operating on the hearts of small children, C. Walter Lillehei occasionally found that his sutures damaged conduction cells in the heart and produced a postsurgical heart block which killed some of his young patients. Lillehei commissioned an external pulse generator, devised by Earl Baaken, a Medtronics engineer, for use in his patients as a temporary assistance device he hoped would allow these conductive cells to heal naturally. When a myocardial pacing wire was inserted into the heart wall and connected to the external device, Lillehei discovered that "one or two volts drove the heart beautifully." The wire lead could be removed without further surgery. Active children required a portable unit, and dependence on the hospital's electrical system inspired the creation of a battery-powered portable unit. The Medtronic 5800, invented in 1958, became one of Medtronic's earliest marketing successes.

Shortly after Beck's success using internal defibrillation, Charles Philamore Bailey, a Philadelphia surgeon, performed the first deliberative and successful intracardiac operation on a 24-year-old young woman with severe mitral stenosis in 1948. Within two years, mitral valve surgery was being performed successfully around the world. Also during the 1940s, Charles Hufnagel, began the formidable task of developing a workable prosthetic aortic valve. In September 1952, Hufnagel successfully implanted a prosthetic "ball" valve into a patient at Georgetown University Medical Center. Without a heart–lung machine, Hufnagel could not replace the faulty valve; he could only position the prosthetic as a "check valve" to correct for the effects of the diseased aortic valve. Nonetheless, the accomplishment was impressive as a first step in creating workable replacement valves.

Heart surgery had lagged technologically behind other surgical specialties in the 1940s. There were a few operations to correct specific abnormalities, but patients risked a mortality rate of 50 percent or more. Heart catheterization techniques, in which a small tube could be moved through a large vein and into the heart chambers, allowed steadily improving visions of diseased hearts, and became an ordinary hospital procedure and part of the standard workup for most patients scheduled for heart surgery during this time. In addition to better preoperative understanding of their patients' heart abnormalities, surgeons required time to fix heart problems surgically. Experiments with hypothermia in cardiac surgery proved that a patient's heart could be stopped "cold" and then warmed up to resume its normal rhythm, allowing surgeons precious time to complete more complicated surgical procedures. Hypothermia and a riskier cross-circulation technique using a live donor were the only options for oxygenating blood during surgery until technological advances brought the first heart–lung bypass machines into the operating suite beginning in 1953. Such advances were critical in the subsequent development of artificial replacement heart valves. Cardiac pacing also remained closely associated with open heart surgery throughout the 1950s.

Between 1957 and 1960, at least eight research groups designed and tested fully or partially implantable cardiac pacemakers in humans. One patient who received one of the earliest pacemakers

was still alive and on his twenty-sixth pacemaker in 2000. Between 1961 and 1963, a great number of implanted pacemakers failed and were replaced with improved models. According to one author, Jeffrey:

> "these were inventions in the early stage of product development. The only justification for using them at all with human beings—a compelling one—was that the patients had little chance of survival without some electrical assist for their heartbeats."

Although some nuclear powered pacemakers were implanted in the U.S. in the 1970s, smaller, more powerful lithium batteries and advanced circuitry made newer pacemakers more attractive. Moreover, by 1972, transvenous insertion of small pacemakers under local anesthesia virtually replaced the major chest surgery that had been required for implanting pacemakers in the 1960s. The development of the heart–lung machine made it possible to remove damaged valves and replace them with mechanical prosthetic valves. The first truly successful artificial valve was developed by surgeon Albert Starr and engineer Lowell Edwards and used in a mitral valve replacement in 1960 at the University of Oregon Medical School. Many variations on this design were quickly tested both in the laboratory and in patients using improved materials and streamlined designs. Mechanical valves, however, do present ongoing challenges. The perfect valve would produce no turbulence as blood moved through the valves, would close completely, and would be constructed of a material that discouraged the formation of blood clots. The danger of clot formation required that patients with mechanical valves take lifelong anticoagulants to prevent clots from forming. Although mechanical valves are extremely durable, they can fail, and when they do they often fail catastrophically. In 1968, Viking Bjork, a Swedish professor created a new mechanical valve with a tilting disk design held in place by welded struts. The design was intended to reduce the risk of blood clots, a significant problem with previous implants. Manufactured by the Shiley Company, approximately 85,000 Bjork Shiley valves were sold worldwide between 1978 and 1986. An engineering flaw (strut failure) in hundreds of the valves, however, caused about 250 reported deaths and even more lawsuits. Nonetheless, for long-term use mechanical valves are still used in about 65 percent of the over 60,000 heart valve replacement surgeries every year.

In both the heart valve industry and the heart pacemaker industry, product failures in the late 1960s and early 1970s created concerns among consumers that led to major changes, both within each industry and within government circles. In 1976, Congress enacted the Medical Device Amendments of 1976 to clarify the scope of the Food and Drug Administration's authority over medical devices. Under the 1938 Food, Drug, and Cosmetic Act, the FDA had taken action against many quack medical devices, and had acted on occasion against medical devices under the drug provisions of the 1938 law, but regulatory actions under such conditions were cumbersome, risky, and expensive. Both industry and the FDA wanted clarification over the agency's authority over the burgeoning medical device field. Dr. Ted Cooper, Director of the NIH National Heart and Lung Institute, chaired a committee that surveyed problems in the medical device industry. Concluding that "many of the hazards seem to be related to problems of device design and manufacture," the report cited over 10,000 patient injuries in the medical literature between 1963 and 1969. Signed into law in 1976, the new law governing medical devices classified them according to their perceived risk. Thus Class 3 devices would be the most tightly regulated, while Class 1 products including such "devices" as sterile bandages, would require less stringent regulation. New Class 3 devices were also subject to premarket approval by the FDA: manufacturers had to demonstrate that they were "safe, reliable, and effective" before they could be put on the market. Debate continues over the long-term effects of government regulation on industry innovation in the medical device field, as in the pharmaceutical industry, but it remains clear that medical device firms, including manufacturers of heart valves and of cardiac pacemakers, have resisted any tendency to make valves and pacemakers mere medical commodities. Competing through innovation has kept prices high and competition in these technology driven industries alive and well into the twenty-first century.

See also **Electrocardiogram (ECG); Hearts, Artificial**

SUZANNE WHITE JUNOD

Further Reading

Foote, S.B. *Managing the Medical Arms Race: Innovation and Public Policy in the Medical Device Industry*. University of California Press, Berkeley, 1992.

Fye, W.B. *American Cardiology: The History of a Speciality and its College*. Johns Hopkins University Press, Baltimore, 1996.

Howell, J.D. *Technology in the Hospital: Transforming Patient Care in the Early Twentieth Century*. Johns Hopkins University Press, Baltimore, 1995.
Kirk J. *Machines in Our Hearts: The Cardiac Pacemaker, The Implantable Defibrillator, and American Health Care*. Johns Hopkins University Press, Baltimore, 2001.
Johnson, S.L. *The History of Cardiac Surgery, 1896–1955*. Johns Hopkins University Press, Baltimore, 1978.
Rothman, D.J. *Beginnings Count: The Technological Imperative in American Health Care*. Oxford University Press, New York, 1997.

Cars, *see* **Automobiles**

Catamarans

The first catamarans (boats with two hulls) were derived from the traditional paddled rafts used by natives of Polynesia. These boats, which were essentially two canoes joined with logs, powered by a single sail, provided great stability and allowed them to travel vast distances on the open ocean.

In Europe, the history of catamarans begins with Sir William Petty (1623–1687). Petty studied the hydrodynamics of model hulls and found that a long, slim hull travels more easily through water than a wide one. The ratio of hull length to hull width, now called the "fineness," "slenderness," or "displacement-to-length" ratio, is the key determinant of hydrodynamic performance. Within certain constraints that we will look at below, a slender hull has less drag and creates less of a bow wave, both of which slow down the movement of boats through the water.

The major disadvantage of such a hydrodynamic hull is its instability. A heavy wind or larger sails, rather than providing more speed, would capsize the hull. The accepted way to increase stability in a monohull while keeping it relatively streamlined is with a heavy keel along the bottom of the hull. However, this decreases the power-to-weight ratio of the boat and forces it to float lower in the water, increasing drag. Petty realized that the stability of a wide boat could be combined with the speed of a thin one by connecting two fine hulls together, producing a light, stable sailing platform. Petty had discovered the key advantage of multihull boats—high transverse stability relative to their displacement.

Despite these early advances, catamaran development did not really take off until the late nineteenth century, when the established yacht designer Nathaniel Herreshoff, from Rhode Island, USA, built the Amaryllis, patented in 1876. Herreshoff had designed a bare-bones structure with a huge sail area and two very light, fine hulls, providing an awesome power-to-weight ratio. Despite beating the fleet of New York Yacht Club's Centennial Regatta, the boat was so unusual it was deemed unfair and banned from all future races.

As skepticism began to wane during the twentieth century, designers began to take advantage of new, cheaper, lighter construction techniques. Because catamarans can use their width for stability instead of ballast, they were better placed to benefit from the development of the new breed of light but strong composite materials. Fiber-reinforced plastics (FRPs) began to be used in boat construction in the 1940s. By the 1970s over three quarters of all boats were made from FRPs. The majority of these used glass-reinforced plastic (GRP, or fiber glass), which is light, strong, cheap, and easy to maintain. Carbon fiber (CFRP), a composite material borrowed from the aerospace industry in the 1970s, provided even greater stiffness, strength and lightness, but remains very expensive. Polyester (in the 1950s), and more recently, carbon fiber- and Kevlar-reinforced sails helped to further reduce weight and improve performance. Boat design, which previously relied heavily upon intuition and experience, benefited from the codification of fluid dynamics models and sophisticated CAD (computer-aided design) tools.

Such innovations allowed faster racing catamarans to be developed, and allowed for bigger and more robust boats. In the world of offshore sailing, large catamarans have set and broken countless speed records. These boats are so efficient that they can travel at twice the speed of the wind that fills their sails (when the wind is at right angles to the direction of travel).

The advantages of two hulls (transverse stability and lightness) are clearest with sailing boats, but engine-powered catamarans can also benefit from the increased stability and decreased drag. Increasingly, fast ferries are built as catamarans, giving a wide, flat, and stable cargo platform. These ferries have taken over much of the role once held by hovercrafts and hydrofoils.

Considering the advantages of two hulls, one might assume that the fast, stable catamaran is the ideal shape for all boats. But there are a few disadvantages that tend to limit their applications to fast, high-performance craft. First, narrow hulls and lack of displacement means that catamarans are bad at transporting large, heavy cargo. Second, catamarans are not efficient when traveling slowly, due to the complexities of fluid dynamics. Slender hulls have a greater "wetted surface area," which is a key constituent of drag at low speeds. As such,

catamarans suffer from a "resistance hump," which must be overcome before increases in power lead to large increases in speed. Third, catamarans are only stable while both hulls remain in the water. Once the boat heels beyond a certain angle, it becomes very unstable and will capsize easily. By contrast, a traditional, weighted monohull will experience a strong righting moment right up until it capsizes. Finally, a catamaran encountering large waves undergoes greater and more complex strains on its structure because it has more points of contact with the water. This presents challenges to designers, who want to take advantage of lightweight (but expensive) materials while also building a robust boat.

Multihull boats (including trimarans with, as you might expect, three hulls) have provided designers with new ways to think about and advance sea travel and racing. The combined advantages of speed, lightness, and stability have overcome hydrodynamic constraints that were once thought insurmountable. It remains to be seen whether there is an upper limit on the speed of these vessels, and whether advances in construction techniques can keep pace with the increasing strains on these boats as they get faster. More radical designs, using hydrofoil technology and fixed-wing sails, look set to challenge speed records further into the twenty-first century.

See also **Hovercraft, Hydrofoils, and Hydroplanes**

JACK STILGOE

Further Reading

The Practical Encyclopedia of Sailing: The Complete Guide to Sailing and Racing Dinghies, Catamarans and Cruisers. Lorenz Books, 2001.

Claughton, A. Wellicome, J. and Shenoi, A. *Sailing Yacht Design: Theory*, Prentice Hall, 1999. For an in depth look at general design issues for sailing boats.

Larsson, L. and Eliasson, R. *Principles of Yacht Design*, International Marine Publications, Camden, Maine, 1995.

Smith, C.S. *Design of Marine Structures in Composite Materials*. Chapman & Hall, London, 1989.

White, R. and Wells, M. *Catamaran Racing: For the '90s*. Ram Press, 1992.

Useful Websites

http://www.therace.org. With information about the catamarans involved in this extreme, round-the-world contest.

Cell Culturing, *see* **Tissue Culturing**

Cell Phones, *see* **Mobile (Cell) Telephones**

Ceramic Materials

Ceramics are synthetic, inorganic, nonmetallic solids that generally require high temperatures in their processing, and are used in many industrial applications. Many of them are metal oxides (compounds of metallic elements and oxygen) but others are compounds of metallic elements and carbon, nitrogen, or sulfur. Typically, ceramic materials have favorable engineering properties such as high mechanical strength, chemical durability, hardness, low thermal and electrical conductivity, as well as relatively low density, which make them comparatively light and suitable for application in, for example, automotive engines. (Magnetic ceramics, used in magnetic memory, are described in the entry on Alloys, Magnetic.)

There are mechanical drawbacks that scientists and engineers have successfully reduced during the last few decades: they are brittle and lack ductility, have poor resistance to mechanical and thermal shock, are difficult to machine because of their hardness, and are sensitive to catastrophic failure because of the potential presence of microvoids.

One of the origins of manufacture of engineering ceramics is in techniques for firing clay for porcelain, tableware, and construction materials. However, twentieth century technology and the chemical, electrical, electronic, automotive, aerospace, medical, and nuclear industries have made new demands on high-performance materials. In the electrical industry, for example, ceramic insulators have been used in electric power lines since the 1900s and since the 1920s in spark plug insulators.

Naturally occurring steatites, a class of magnesium silicate minerals also known as soapstone, were known and used in the late nineteenth century for burners in stoves and gas lights, but were developed and processed as improved electrical insulators that had low loss at high frequencies and high temperatures. They were therefore used as insulators in microwave electronics for radio relays in telephone networks in the 1940s. During and after World War II the applications of ceramics in electronics extended to compact capacitors, piezoelectric transducers for use in telecommunications, resistor and semiconductor compositors as well as magnetic materials and other energy converters.

Apart from electric and electronic applications abrasives have been a field in which ceramic materials, in this case silicon carbide (SiC) and aluminum oxide (Al_2O_3), excelled. During the nineteenth century it became clear that abrasive products like natural sandstone used in grinding

wheels no longer satisfied industrial demands. In 1891 Edward G. Acheson, an electrical engineer from Monongahela City in Pennsylvania, combined a mixture of clay and powdered coke in an electrical furnace. This resulted in shiny crystals (silicon carbide, SiC) which proved immensely suitable for polishing precious stone. Until the invention of boron carbides in 1928, silicon carbide had been the hardest synthetic material available. Silicon carbide's high thermal conductivity, strength at high temperatures, low thermal expansion and resistance to chemical reaction made it the material of choice for manufacturing bricks and other refractories at high temperatures, for example for industrial boilers and furnaces, and tiles covering the space shuttles.

The manufacture of aluminum oxide abrasives closely followed the development of silicon carbide. In 1897 scientists at the Ampere Electro-Chemical Company, New Jersey, made the first successful attempts at aluminum oxide manufacture with rock bauxite, of which aluminum oxide is the main ingredient. Today aluminum oxide, sometimes with the addition of zirconium oxide, is indispensable for producing highly precise and ultrasmooth surfaces in the automotive and aerospace industries.

During the twentieth century, high-performance engineering materials were increasingly required for structural applications. In particularly erosive, corrosive, or high-temperature environments, materials such as metals, polymers or composites could not fulfill the demands made on them. In 1893 Emil Capitaine, a German engineer who played a role in the development of the internal combustion engine, suggested the use of porcelain and fire clay in engines, though it is not clear what for; a decade later engineers at the Deutz Motor Company in Cologne experimented with ceramic materials for use in stationary gas turbine blades. During World War II interest in new high-performance materials grew rapidly, in Germany not so much for surpassing steel alloys but for replacing metals such as chromium, nickel, and molybdenum, which were in short supply in the armament industry. Experiments with aluminum oxide seemed promising for use in gas turbine blades but the material's susceptibility to thermal shock created insurmountable difficulties. In order to reduce these problems and to impart greater ductility and thermal shock resistance to aluminum oxide, engineers during and after World War II tried to mix ceramics with different metals to make composite materials. Although this did not yield the expected results at the time, the idea of reinforcing a ceramic matrix with metal fibers proved useful. Metal components have also been made stronger by the incorporation of ceramic fibers, whiskers, or platelets; special ceramic fibers and whiskers are also incorporated in a ceramic matrix to increase toughness and reduce the risk of catastrophic failure due to incipient cracks and microvoids. Automobiles and aircraft engines of the future may have ceramic matrix composite components such as brake disks, and turbine parts for high-temperature jet engines.

After the war, it soon became clear that for many applications a ceramic material like aluminum oxide had too many drawbacks, a conclusion based largely on research in Germany. Attention therefore shifted from oxide ceramics to ceramic materials such as silicon nitride or silicon carbide. During the 1970s the oil price shock accelerated interest in thermal efficiency of power plants, enhanced by environmental legislation in countries such as the U.S. or Germany. If the combustion temperature is raised for greater efficiency, the turbine parts must be oxidation-, impact-, and thermal-shock-resistant. From that time onward, government-sponsored research and development programs, especially in the U.S., Japan, Britain, France and Germany, have advanced research on ceramic materials such as silicon nitride and silicon carbide which, among other assets, are more oxidation resistant than super alloys. To date, however, none of these ceramic materials has been successfully adapted to the proposed high-temperature gas turbines. Advanced structural ceramics have also been employed in nuclear power as heat-resistant control rods.

Because of their good biocompatibility, ceramics are employed in medical and dental applications such as false teeth, implants, and joint replacements; in the automotive industry they are useful as catalysts, catalyst supports or sensors.

Apart from silicon nitride and silicon carbide, zirconium oxide is an excellent engineering ceramic. Zirconium oxide offers chemical and corrosion resistance at far higher temperatures than aluminum oxide. Stabilized zirconium oxide produced by addition of calcium, magnesium, or yttrium oxides, exhibits particularly high strength and toughness. This and its low thermal conductivity led to applications in oxygen sensors and high-temperature fuel cells. Over recent decades it has become possible to better cope with the intricacies of engineering ceramic materials but many questions are still unsolved. The ceramics' first-rate potential in various demanding applications make these efforts worthwhile.

See also **Alloys, Magnetic**

HANS-JOACHIM BRAUN

Further Reading

Braun, H.-J. Engineering ceramics: research, development and applications from the 1930s to the early 1980s, in *Science–Technology Relationships*, Herlea, A., Ed. San Francisco Press, San Francisco, 1993, pp. 161–167.

Braun, H.-J. and Herlea, A., Eds. *Materials: Research, Development and Applications*. Brepols, Turnhout, Belgium, 2002.

Brook, R.J., Ed. *Concise Encyclopaedia of Advanced Ceramic Materials*. Pergamon Press, Oxford, 1991.

Kingery, D.W.D., Ed. *High Technology Ceramics. Past, Present, Future: The Nature of Innovation and Change in Ceramic Technology*. American Ceramic Society, Westerville, 1986.

Mason, T.D. Industrial ceramics, in *The New Encyclopaedia Britannica, Macropaedia*, 15th edn, vol. 21. Encyclopaedia Britannica, Chicago, 2003, pp. 246–265.

Richerson, D.W. and Dunbar, B.J. *The Magic of Ceramics*. American Ceramic Society, Westerville, 2000.

Changing Nature of Work

Two men, Fredrick Winslow Taylor (1856–1915) and Henry Ford (1863–1947), are often credited with transforming the relationship between humans and machinery in the workplace and beyond, lending their names respectively to overlapping twentieth century socio-technical systems, Taylorism and Fordism. Subsequent developments in production technologies like automation, flexible specialization, and "lean production" were compelled to engage with their enduring legacy.

Taylorism

Taylor's *Principles of Scientific Management* (1911) set out a methodical engineering logic for eliminating wasteful human effort. In the famous "Schmidt" experiment Taylor turned his close observations of "inefficient" working practices and his training in engineering exactitude to determine in minute detail only those motions essential to the completion of a defined work task, in this case the handling of pig iron. Taylor thus answered his own rhetorical question: "Why can we not apply the same principles of efficiency to the hand and muscle of man that we apply to the design of machines?" External measurement and control could manipulate human activity to make it more "machine-like." Thus the technical *conception* of work design was radically separated from the laboring *execution* of work activity. Job analysis and time and motion study allowed work previously done by a single person to be broken down into more simple, repetitive tasks. Pay rates were calculated more precisely according to relative time spent handling work or merely supervising the machine at work. Large gains in productivity seemed to support Taylor's insistence that the technical organization of work be standardized and synchronized to allow human labor to perform efficiently.

Many objections have been aimed at the Taylorist rationalization of work. One is that Taylor had a very crude understanding of worker motivation as animated by economic progress. He anticipated that the elimination of waste, the removal of complicated decision making for the worker through the use of scientific procedures and piecework pay schemes would appeal greatly to a basic human desire to improve their own material well-being. A second objection, made by radical and Marxist critics like Harry Braverman (1974), is that Taylorism was far from being a scientifically disinterested exercise. Instead, Taylorism merely provided a rationalization for passing control of the production process away from labor and into the hands of technicians and supervisors who performed functions essential to the class interests of employers. A third objection raised by psychologists working in the behavioral science tradition is that Taylorism was not scientific enough. It lacked scientific rigor in human physical or psychological capacities and was based on arbitrary decisions and informed guesswork about the mind–body coordination of "average" or "experienced" workers. A further objection is that workers are not merely passive factors of production animated by the single viewpoint of a management-designed socio-technical work system. Workers can actively modify or subvert even highly detailed and closely monitored work systems like scientific management, both informally as members of a work group and more formally through trade unionism. Finally, though Taylorist ideas became widely diffused, especially under the impact of the production demands of two world wars, they were rarely adopted as an entire scientific system but were pragmatically deployed by engineers and managers moved job planning from the shop floor into the offices of production engineering departments.

Fordism

With the development in 1908 of the Model T automobile Henry Ford invented what he himself

called "mass production." Others like Antonio Gramsci gave it the eponymous title "Fordism" to indicate the ways that mass production and mass consumption brought an entirely new way of life into existence. In his contribution to the thirteenth edition of the *Encyclopaedia Britannica* on "Mass Production," Ford set out his new production philosophy—simplicity.

> "Three plain principles underlie it: (a) the planned orderly and continuous progression of the commodity through the shop; (b) the delivery of work instead of leaving it to the workman's initiative to find it; (c) an analysis of operations into their constituent parts."

These principles mechanically controlled the pace at which work passed before the worker. From Ford's 1913 experiment with a conveyor belt for the assembly of magnetos at the Highland Park plant in Detroit the "three plain principles" were generalized to the assembly of other automobile components and, in the 1920s, to a broad range of consumer goods. Huge gains in productivity meant that small batch goods previously limited to the luxury market became mass produced in standard forms for relatively low prices. Inside the car plant Ford's real achievement may have been less to do with the continuous pace of the work than with the simplicity of easily assembled, inexpensive, perfectly interchangeable parts fashioned to high tolerances and amenable to the same gauging system.

On the consumption side, Ford's success was based on low, and falling, costs and durable and easily serviced car design and materials. However, Ford's initial advantages slipped in the 1920s when Alfred P Sloan's new functional management system at General Motors adopted a thorough standardization of mechanical items coupled with annual design changes to the external appearance of the car's "hang-on" features. Yet, for all Ford and Sloan's innovations, autoworkers were viewed as interchangeable parts of the production process, something rendered artistically in the controversial 1932 Detroit Industry frescoes painted by the radical Mexican muralist Diego Rivera. To offset labor discontent, in 1914 Ford offered workers "the $5 day," effectively doubling wages for those workers that qualified and created a "Sociological Department" to enforce respectable social behavior among the workforce. As turnover at car plants decreased workers became increasingly dissatisfied with their experiences of working on the assembly line and car plants became centers of industrial conflict and unionization drives.

Automation

Automated machine tools took the process of component standardization, interchangeability and technical control a stage further. Automation provided a managerial solution to the high-cost burden and inflexibility of jigs and fixtures, which left high levels of control and discretion in the hands of toolmakers. Machine tools like center lathes or milling machines are indispensable to all machine-based metal-working processes, including the capital goods production of machine tools themselves. Machine tools rotate or otherwise set in motion the tool piece or the work piece (or both) to reshape the work piece to a predetermined size, contour, and surface finish.

In the 1930s and 1940s, tracer technology increasingly deployed hydraulic or electronic sensing devices to cut contours into materials according to templates. More sophisticated sensing and measuring devices like precision servometers were only one example of industrial innovation during World War II. The challenge was how to completely automate machine tools. Unlike automated manufacturing processes, where single-purpose, fixed automation proves cost effective for high volume output, automated machine tools are more versatile, multipurpose machines geared towards one-off or small-batch volumes. Automated machine tools combine machine versatility with off-the-job control and information records. One solution was "record-playback," developed around 1946 by General Electric and Gisholt, and some smaller companies. A record was made on magnetic tape of a machine set in motion under the instruction of a machinist, allowing the automatic production of identical parts by replaying the tape and precisely reproducing the machine tool movements.

A second solution, numerical control (N/C), was developed by John Parsons, a Michigan subcontractor who manufactured rotor blades for U.S. Air Force helicopters. N/C involved what David Noble (1984) calls "an entirely different philosophy of manufacturing." N/C transferred the knowledge of the machinist to programmers who made up the tapes prior to the planned work arriving at the machine. Early programming involved manually writing unique subroutines for each particular geometric surface, a time-consuming and tedious operation. In 1956 Douglas Ross, a young engineer and mathematician at MIT, developed a systematized solution, automatically programmed tools (APT) that combined the versatility of five-axis control with an overarching skeleton computer program. The U.S. Air Force funded N/C research

and enforced its diffusion among its contractors, whose cost-plus military contracts subsidized their adoption of the expensive APT system, despite the operational difficulties caused by the use of large, complex computing systems.

N/C was viewed enthusiastically by engineers, managers, and the Air Force as a hyper-Taylorist solution to human error and recalcitrant worker attitudes. As Peter Drucker stated, "automation is a logical extension of Taylor's scientific management" (in Kranzberg and Pursell, 1967). However, controlling labor through automatic machinery was far from complete since the implementation of such complex, expensive and fallible equipment was contested by the machinist union and a continuing need for highly skilled machinists came to be recognized by many manufacturers. N/C techniques were further enhanced by advances in electronics such as computer-aided design and manufacture (CAD/CAM), which seemed to threaten even the skills of design engineers. Such arguments tend to rely on technological determinism, where technology is viewed as an independent force shaping social relations rather than being shaped by them. CAD/CAM has, however, had the support of engineers, reflecting their senior social positions inside the hierarchy of command and the division of labor.

Women and Industrial Technology

Until the final quarter of the twentieth century the relationship between gender and technical change was subsumed under the simple equation of masculinity with machinery. The prevailing view tended to see technology as neutral and that men's physical strength equipped them to adopt machinery more readily than women. Such biological reasoning failed to account for the social shaping of technology by existing ideologies of gendered work. After all, machinery tends to be introduced to reduce physical effort in the production process. Technical work often fosters a struggle for control of nature in 'raw materials' through the controlled use of energy, something many feminist writers view as intrinsically male attributes.

Women are subject to systematic discrimination in pay and access to skilled occupations, even in industries such as textiles where women predominate. Women also tend to be regarded, falsely, as transient or peripheral to the core of the labor market and passive in the face of employer controls. Indeed, the labor force has been increasingly feminized, partly due to the expansion of services and part-time work. This reflects a domestic division of labor that casts women as "carers" in the family. The ideology of women as home makers has also had profound consequences for the way industrial technologies were introduced into the domestic sphere. Women's unpaid domestic labor was subject to an invisible industrial revolution in the twentieth century—from hand power to electric power, to gas and electric appliances for cooking, cleaning, waste disposal, to central heating, to prefabricated housing, and so on.

Industries that employ relatively low-cost and abundant female labor face little incentive to invest in expensive labor-saving technical innovation. Automation is subject to technical as well as social limits. For example, the skilled hand–eye coordination of female sewing machinists has proven difficult to fully substitute with technology. Industries such as clothing traditionally deploy a highly defined sexual division of labor, with men designated the high status of "tailor" and women the lower status of "seamstress/" As technology entered the industry craftsman tailors abandoned machinery to women. Men retained craft status in the cutting room until cutting and pattern process were separated and routinized by the advent of Taylorism and male craft workers were substituted by deskilled female workers.

Studies of the introduction of word processing show how office technologies such as typewriters, dictation machines, copiers, printers, and filing cabinets, over which women workers previously exerted some control and autonomy, were replaced by the "intelligent office" of networked word and data processing through advanced computer and telecommunication systems. A loss of status and control accompanies female secretaries and typists who previously typed their copy but now simply key-in data and follow screen prompts. Tasks such printing, filing, and typing become fragmented and physical movement is reduced by systems that store the data electronically.

The Rise of Services

The socio-technical system of simplifying and replacing labor with machinery set in motion by Taylor and Ford and deepened by process automation began to falter in the developed economies in the1960s. On the other hand, state policies in support of industrial imitation allowed newly industrializing countries (NICs) such as Taiwan and South Korea to successfully industrialize. A new international division of labor began to emerge as multinational corporations were able

to take advantage of the division and subdivision of production and advanced transport and communication techniques to relocate labor-intensive subprocesses to low-cost Third World locations. As manufacturing was subdivided internationally, industrial employment went into absolute decline in Western Europe and the U.S. The Western adoption of neoliberal principles in the 1980s helped further stimulate what many described as "globalization." This did not mean that the entire world became fully integrated. Instead, productive investment strategies created a triad of regionally linked advanced economies, North America, Europe, and South East Asia, exacerbating the development gap between the triad and the NICs, and the low-industrialized countries making up the Third World.

Accompanying the deindustrialization of Western economies was a continuous rise in the absolute numbers employed in the service sector. However, even with fewer workers manufacturing output continued to rise, indicating manufacturing's socio-technical capacity for increasing productivity per worker. Due to its heterogeneous nature, "services" are notoriously difficult to define. *The Economist* defined services as "anything sold in trade that could not be dropped on your foot." Services are often viewed as some kind of residual category, what is left over after manufacturing, agriculture, and extractive sectors are accounted for. Yet services now appear to be central to most developed economies. Services can be broken down into two basic concepts:

1. Service *industries*, where the final output takes a nonmaterial form such as financial services
2. Service *occupations*, which includes jobs not directly involved in producing material goods such as transport, sales, clerical, administrative, professional, and cleaning services.

Postindustrialism or Industrial Services?

The rise of service industries and service occupations has been accompanied by much dispute over how to characterize changes to society and economy. For some, like Daniel Bell in the U.S. and Alain Touraine in France, "industrial society" is being superseded by "postindustrial society." Bell describes a trajectory of socio-economic development, which passes through preindustrial, industrial, to postindustrial society. Preindustrial society was technologically limited to a brute human struggle with nature based on raw, muscle power." Industrial society was premised on energy-driven sectors that coordinated men and machines for the mass production of goods. Postindustrial society is

> "organized around knowledge, for the purpose of social control and the directing of innovation and change" [*Bell, 1973*].

On Bell's calculations, by 1970, 65 percent of the U.S. labor force worked in services such as information and professional services, health, education, welfare, cultural, and leisure services and research, design and development.

Touraine similarly identifies a shift from manufacture-based industrial society to a knowledge-based postindustrial society. Where Bell draws conservative conclusions about social and political development, Touraine identifies new, radical forms of protest. In industrial society the workplace became the organizing center for social movements based on economic class which appealed to the national state for redress of grievances. In postindustrial society, new social movements like ecology and antinuclear movements emerge that reject technocratic integration into the economy and the state. This reflects the more diffuse sources of political power in a knowledge-society characterized by Touraine as the "management of the data-processing apparatus."

Both Bell and Touraine make their claims about postindustrial society by placing a conceptual emphasis on knowledge." Yet the relationship between knowledge-based services and physical goods manufacturing is more complex than their image of separate, contiguous sectors allows for. Indeed much of the rise in services can be accounted for by the greater support necessary for advanced technical production systems. Additionally, an increased dependence on knowledge may still rely heavily on production in the capital goods sector of intelligent machines and knowledge infrastructures.

Gershuny and Miles identify the emergence of a "self-service" economy that defies the linear logic of the supplanting of industrial society by postindustrial society." For them *service functions* can be met by a range of organizational and technical forms. Since services tend to be more labor-intensive and less amenable to large productivity gains arising from technical innovation, households can elect to service themselves by substituting cheaper manufactured goods such as cars, washing machines, videos, microwave ovens, chilled ready-made meals, and ready-to-assemble furniture, for market-based services such as transport, laundries,

cinema, restaurants, and so on. The self-service economy therefore represents a further extension of manufacturing into everyday life rather than its eclipse by "knowledge." The rise of the self-service economy also provides a good example whereby a range of technical innovations in the manufacture of consumer goods cluster in the 1950s to reinforce what Joseph Schumpeter referred to as a "wave of innovation." Schumpeter's model of discontinuous, disruptive periods of technical change contrasts with the linear, continuous model of socio-technical succession in post-industrial theories.

After Fordism

Fordism reached a high point in the 1950s during the postwar economic boom. Fordism was never an all-embracing socio-technical system and the term does not cover the wide variety and processes even in the manufacturing sector. At best, Fordism represented the leading edge of development and is not an accurate description of the entire structure of production between the 1920s and 1970s.

Two main schools of thought emerged to characterize the period after Fordist industrial leadership. First, Michael Piore and Charles Sabel in *The Second Industrial Divide* (1984) deployed an institutional approach to technological development which focused on the contingent influence of particular national political and economic differences. They discuss the crisis of Fordism in terms of external shocks such as labor discontent, and the 1973 oil crisis for triggering inflation and recession which repressed demand and investment and led to the fragmentation of mass markets. This trend was exacerbated by the internal structural tensions of mass production where the rigidities of Fordist production could not adapt to the saturation of the mass market and consumer demands for more differentiated and individualized products. Piore and Sabel advocated the adoption of "flexible specialization" as a solution to the crisis of Fordism. Flexible specialization deploys newly versatile production technologies based on computerized systems to efficiently switch production from high volume mass production to small batches of customized goods.

In the work of the French Marxists Michel Aglietta, Robert Boyer, and Alain Lipietz, a second approach focused on "neo-Fordism" as the latest stage in capitalism's development. Fordism enters crisis in this account because of the contradiction between the "regime of accumulation,"; that is, the macroeconomic principles around which capitalism stabilizes for a period of time, and "the mode of regulation,"; that is, how a particular national economy is regulated both within and between countries. After the crisis of overproduction in the 1930s, the Fordist regime of accumulation was characterized by mass production where real wages rose in line with productivity, a stable rate of profit, and full employment. This was buttressed by a new mode of regulation consisting of institutionalized collective bargaining, a developed welfare state, and the control of credit by central banks. Neo-Fordism is characterized by flexible production technologies allied to differentiated and segmented consumption, and a less comprehensive welfare state.

Where the institutionalists focus on the contingent national factors, the regulationalists focus on the structural contradictions of whatever manifestation capitalism takes. Both approaches have been criticized for empirically neglecting the resilience of mass production and mass consumption. Taylorism has deepened into other areas such as the mass production and consumption of fast foods, giving rise to the further rationalization of all social institutions in what George Ritzer called the "McDonaldization" of society. Both have also been criticized for conceptually isolating or reifying a particular production technology and erecting an elaborate theory of socio-economic and cultural change around it.

The Coming of the Information Age

Multinational companies in Europe and the North America also began to adopt "Toyotism," or "lean production," after the Japanese manufacturing philosophy of Taiichi Ohno. Toyotism challenged the top-down controls over labor of Taylorism by a system of workers' own continual improvements to the production process called *kaizen*. All the stages in the creation and circulation of the commodity were to be synthesized through the *kan-ban* system by developing a long-term relationship between the manufacturer, the dealer, and the consumer through a "build-to-order" system rather than the unpredictable "build-to-economize" system of mass production, which regularly resulted in crises of overproduction. On the other hand, the flexible worker needed to become multiskilled to cope with the new processes and technologies.

New technologies based on information processing systems are transforming certain aspects of work, though the extent of this should not be exaggerated. In the early 1980s the Fiat car factory at Cassino claimed to be the first fully robotized

plant in Europe. Teleworking, for instance, was claimed to herald a new era of social organization, where the home would increasingly absorb the functions of the workplace through being networked to the Internet. In the late 1990s the available evidence indicated that only very small numbers "tele-commute" with little daily reduction in car use, although the myth of workplace transcendence and travel reduction remains strong. Information and communication technologies are also being used to stretch and intensify the working day, with the average office worker estimated to be spending three hours per day sending and receiving some 150 electronic messages. Strict control over and close monitoring of labor is even more evident in technologically intensive workplaces such as call centers, which proliferated on the edge of low-cost, metropolitan centers in the 1990s. Call centers utilize automated call distribution systems to direct incoming or outgoing calls to customer service operators, housed in large sheds that resemble advanced telecommunications factories.

See also **Automobiles; Farming, Agricultural Methods; Farming, Mechanization; Iron and Steel Manufacture; Technology and Leisure; Urban Transportation**

ALEX LAW

Further Reading

Aglietta, M. *A Theory of Capitalist Regulation: The US Experience*. New Left Books, London, 1979.

Baker, E.F. *Technology and Women's Work*. Columbia University Press, New York, 1964.

Bell, D. *The Coming of Post-Industrial Society*. Heinemann, London, 1973.

Braverman, H. *Labor and Monopoly Capital*. Monthly Review, New York, 1974.

Cavendish, R. *Women on the Line*. Routledge & Kegan Paul, London, 1982.

Cockburn, C. *Machinery of Dominance: Women, Men and Technical Know-How*. Pluto Press, London, 1985.

Edwards, R. *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. Heinemann, London, 1979.

Flink, J.J. *The Car Culture*. MIT Press, Cambridge, MA, 1979.

Ford, H. Mass production, in *Encyclopaedia Britannica*, 13th edn, suppl. vol. 2, 1926, pp. 821-823.

Gershuny, J.I. and Miles, I.D. *The New Service Economy: The Transformation of Employment in Industrial Societies*. Macmillan, London, 1983.

Gillespie, A. and Richardson, R. Teleworking and the city: myths of workplace transcendence and travel reduction, in *Cities in the Telecommunications Age: The Fracturing of Geographies*, Wheeler, J., Aoyama, Y. and Wharf, B., Eds. Routledge, London, 2000.

Gorz, A. *Reclaiming Work: Beyond the Wage-Labour Society*. Polity Press, Cambridge, 1999.

Gramsci, A. Americanism and Fordism, in *Selections from the Prison Notebooks*, Hore, Q. and Nowell-Smith, G., Eds. Lawrence & Wishart, London, 1971.

Greenbaum, J. *Windows on the Workplace: Computers, Jobs and the Organization of Office Work in the Late Twentieth Century*. Monthly Review Press, New York, 1995.

Hirst, P. and Thompson, G. *Globalisation In Question*. Polity Press, Cambridge, 1996.

Hounshell, D. *From the American System to Mass Production 1800–1932*. John Hopkins University Press, Baltimore, 1984.

Jaffe, A.J. and Froomkin, J. *Technology and Jobs: Automation in Perspective*. Fredrick A Praeger, New York, 1968.

Knights, D and Willmott, H., Eds. *Gender and the Labour Process*. Gower, Aldershot, 1986.

Kranzberg, M. and Pursell, C.W., Eds. *Technology in Western Civilization, Volume II: Technology in the Twentieth Century*. Oxford University Press, New York, 1967.

Lacey, R. *Ford: The Men and the Machine*. Heinemann, London, 1986.

Lipietz, A. *Mirages and Miracles: The Crisis of Global Fordism*. Verso, London, 1987.

Littler, C.R. *The Development of the Labour Process in Capitalist Societies*. Heinemann, London, 1982.

Noble, D. *Forces of Production: A Social History of Industrial Automation*. Alfred A Knopf, New York, 1984.

Pollard, S. *The Genesis of Modern Management*. Edward Arnold, London, 1965.

Pollert, A. *Girls, Wives, Factory Lives*. Macmillan, London, 1981.

Ritzer, G. *The McDonaldization of Society*. Sage, London, 2000.

Schumpeter, J. *The Theory of Economic Development*. Harvard University Press, Cambridge, 1934.

Smith, C. *Technical Workers: Class, Labour and Trade Unionism*. Macmillan, London, 1987.

Taylor, F.W. *Scientific Management*. Harper & Row, New York, 1947.

Taylor, P. and Bain, P. An assembly line in the head: the call centre labour process. *Ind. Relat. J.*, 30, 2, 101–117, 1998.

Thompson, P. *The Nature of Work: an Introduction to Debates on the Labour Process*. Macmillan, London, 1989.

Womack, J.P., Jones, D.T. and Roos, D. *The Machine That Changed the World*. Macmillan, New York, 1990.

Wood, S., Ed. *The Degradation of Work?* Hutchison, London, 1982.

Chemical Process Engineering

The chemical industry expanded dramatically during the twentieth century to become a highly integrated and increasingly influential contributor to the international economy. Its products seeded and fertilized the growth of other new technologies, particularly in the textiles, explosives, transport, and pharmaceutical industries. The industry also became a major supporter of industrial

research, especially in the U.S. and Germany. The production of chemicals during the century can be described as a history of products, processes and professions.

At the turn of the twentieth century, the chemical industry was dominated by two classes of products. While a variety of chemicals had been produced on a small scale since antiquity, the nineteenth century had seen the rise of large-scale chemical industries. The manufacture of bulk (or heavy) inorganic chemicals such as sulfuric acid and bleaching powder became important industries by midcentury. Coal-tar derivatives were another important category of product. The production of illuminating gas from coal from the 1820s generated significant waste materials which subsequently were processed into useful products in their own right; for example, naphtha, used as lamp fuel and solvent; creosote, for the protection of timber (especially railway sleepers); and dyes. Synthetic dyes, originating with the discovery in 1856 of mauve from a coal-tar derivative by the Briton W.H. Perkin, became an industry increasingly dominated by Germany from the 1870s; by World War I, Germany was producing 80 percent of the world output of dyestuffs and other fine chemicals reliant on complex processes. By the mid-twentieth century, these products had been eclipsed by spectacular increases in the production of petroleum, petrochemicals, synthetics and pharmaceuticals. Nevertheless, the industry remained little appreciated by the public because most of its output was in the form of intermediate products for use by other industries.

The larger scale changes in chemical production can be better understood in terms of processes rather than as discrete products. Indeed, Hardie and Pratt (1966) describe the history of the chemical industry in terms of the history of its processes; that is, the succession of actions that transform raw materials into a new chemical product. Such conversion may involve chemical reactions (e.g., the production of soda alkali and sulfuric acid in the LeBlanc process); physical change (e.g., oxidation by roasting, or distillation by boiling and condensation); or physical manipulation (e.g., by grinding, mixing and extruding).

Professions, too, had an important influence on the trajectory of the industry. Two models, developing side by side, supplied the design and production labor of the chemical industries in different countries. At the turn of the century, chemists held sway over the technical side of much of the business of chemical manufacturing. But the growing economic importance of bulk chemicals over the previous few decades had been accompanied by the emergence of new specialists in the design and operation of chemical plants. From the 1880s certain groups of technical experts dissatisfied with their standing in the industrial hierarchy made repeated attempts to wrest control of certain industrial practices from chemists. By the first two decades of the twentieth century, groups of senior workers employed in American and British chemical manufacturing were seeking to distinguish themselves from analytical chemists, who had a low status, by claiming as their own the task of scaling-up chemical processes from the laboratory to the industrial level. In America and Britain, a new breed of worker—the chemical engineer—developed as embodying this special expertise. The growth of a "process" perspective developed in lock-step with this nascent profession. This culminated in the formation of the American Institute of Chemical Engineers (1908) and, in Britain, the Institution of Chemical Engineers (1922).

The leaders argued that the development of new manufacturing processes could be understood within the conceptual framework of industrial chemistry. Early chemical engineers used their knowledge of physical chemistry as a way of distinguishing themselves both from self-trained "factory hands" and from mechanical engineers who had become involved with chemical plant design.

The key to legitimating chemical engineering as a profession was the evolution of a new cognitive base—the unit operation. Rudiments of the idea can be found in the writings of George E. Davis, a Manchester chemical engineer, from the late 1880s (Davis, 1901). The unit operations evolved more explicitly at several American universities, especially the Massachusetts Institute of Technology. The unit operations were first described as such in 1915 by Arthur D. Little, who defined them as discrete physical processes employed in chemical manufacturing, such as distillation, roasting, filtering, and condensation. Each described a particular way in which material could be transformed physically—for example, by the reduction in size of solid matter, or by the mixing or separation of solids, liquids, or gases. These basic processes would be performed in sequence to obtain a final product. While the number and order might vary from chemical to chemical, any manufacturing process could be understood in terms of the same set of building blocks. The steps could either be carried out on batches of material, or performed as a continuous process in which the material is transformed at different locations in a plant. This

intellectual transformation made possible a scaling up of production and greater efficiency in an industry that had blighted the landscape by serious air and water pollution from its waste products during the nineteenth century.

The particular technical requirements of petroleum refining—namely physical operations such as distillation—may explain the acceptance of the conceptual framework of the unit operations by manufacturers in America, where oil had been pumped since the late nineteenth century. Hence the American profession appears to have been shaped by at least three factors: the desire of certain groups of technical workers to differentiate themselves from the more heterogeneous and less prestigious category of "chemists"; their appropriation of a cognitive realm (first theoretical chemistry, and then unit operations) as a way of underpinning claims to technical expertise; and a gain in legitimacy in academic, engineering and industrial circles following the successful application of this intellectual apparatus to the design and operation of large chemical plants.

By contrast to the American case, a unique occupation combining mechanical and chemical expertise failed to coalesce in Germany, Austria and Switzerland, major chemical producers before and during the twentieth century. In the German occupational model, labor was divided strictly between industrial chemists and mechanical engineers working in tandem. One reason for this difference is that the petrochemical industry there was negligible, and much more complex chemical syntheses—traditionally based on low-volume batch processes and requiring the specialized scientific knowledge of chemists—dominated the dyestuffs and pharmaceuticals industries. The occupational specialty of *Verfahrenstechnik* became organized around specific industrial products and their manufacturing processes, and chemical plants through most of the twentieth century were designed and maintained by a combination of chemists and mechanical engineers (Buchholz, 1979).

In Britain, the chemical engineering perspective gained ground during the First World War. The organizer of government chemicals factories, Kenneth B. Quinan, was credited with introducing clear methods of analyzing the problems of chemical design and production. Quinan's techniques of statistical control for administration achieved impressive manufacturing efficiencies, and were important in helping to demonstrate that chemical engineers were competent enough to handle the problems of rapidly and efficiently scaling-up facilities from laboratory demonstration to "pilot plant" to "production plant" size. British chemical engineers stressed the importance of industry-wide research, academic training, and accreditation to carry forward the expansion triggered by the wartime chemical factories. While promoting similar professional and educational aims as their American counterparts, they also emphasized government–industry–professional cooperation (so-called "corporatism") and the efficient recovery and use or disposal of byproducts.

World War I had a catalyzing effect in ramping up chemical production and in defining the specialists responsible for it. It also played an important role in the coalescence of representative bodies that came to promote the industry, discipline and profession. The supply of chemical munitions was central to the waging of the war. At the outbreak of World War I, only the German chemical industry had been prepared for the required scale of explosives production; high explosives, in particular, were produced by few factories. For other materials such as cordite (the principal propellant used by the British military, and consisting principally of a mixture of nitroglycerin and nitro-cellulose extruded in thick cords), there appeared to be no prospect of a large-scale demand after the war which would induce existing manufacturers to extend their works. Thus in Britain, the government became directly involved in the wartime management of chemical production. In the U.S., the Du Pont company was subsidized by the government to increase production. The "chemists' war" eventually mobilized thousands of technical workers with chemical and engineering backgrounds.

These wartime experiences dramatically altered postwar chemical production. Chemical firms—traditionally secretive about the processes they employed—were now larger and less isolated. While some postwar factories were sold for little more than scrap, other explosives plants were readily converted to civilian products such as rayon production, for which the production and raw materials were not dissimilar to gunpowder production. To many such transformed plants came experienced managers and designers from the ordnance factories.

This mixing of experience in the commercial and government spheres, and the heightened international competition after the war, encouraged firms to combine and expand. In the U.S., the DuPont company, which had begun as an explosives manufacturer, expanded into dyestuffs and cellulose products during and after World War I. In

Germany, IG Farben amalgamated in 1925 from dyestuff firms that had been associated since 1916. As the largest chemical company in the world, it continued to be the most prolific inventor of new synthetic materials. In Britain, Imperial Chemical Industries (ICI) was formed in 1926 by the merger of a number of companies, and the Association of British Chemical Manufacturers was formed to unite industry interests for the first time. In France, RhÔne-Poulenc was created in 1928 by the merger of two fine-chemical firms.

New technologies, too, signaled new opportunities between the wars. The most prominent of these was high-pressure plant, pioneered in Germany by the Haber–Bosch process for the synthesis of ammonia (1909). This demanding technology, employing gases at high pressure and temperature, was inherently large-scale and capital-intensive. The plant required expertise in compressors, pumps, reacting vessels, control devices and moving machinery. These demanding technologies were emblematic of a new scale and approach to process engineering.

New materials for chemical plant were also a frequent source of enthusiasm and optimistic forecasts. Novel metal alloys and corrosion-resistant materials had been sinks for speculative capital for decades, and proliferated with somewhat more success during the interwar period. Developments in tantalum, tellurium, lead, copper, and nickel alloys transformed the environment of chemical plants. Stainless steel, first used in German and Austrian nitric acid and munitions plants after 1916, became increasingly applied from the 1920s in an industry which had (prewar) relied on cast and wrought iron, mild steel, lead, copper, tin, wood, and bitumen. Chemical stoneware, vitreosil and inert resins similarly multiplied the options available to designers, and improvements in welding techniques allowed stronger, more reliable vessels than could be produced by forging or riveting. Other singular technological innovations became just as pervasive: electric motors, for example, helped promote a transformation in the practice of pumping fluids, which had frequently been based on pneumatically driven lifts and "eggs" before World War I. A postwar flurry of books, periodicals and advertisements supported this new awareness of a coherent chemical process industry.

World War II further scaled up chemical production capacity in many countries. Most factories were organized either as government-owned, company-operated (GOCO) plants in America or as "agency" factories in Britain, in which a firm constructed and operated plants for government production quotas. As with the previous war, there were long-lasting consequences for production capacity, product expertise and commercial relationships.

During the decade and a half after World War II, chemical engineering became firmly established. The chemical and process industries evinced an unprecedented expansion, and chemical engineers found new and more senior places within them. The discipline became firmly rooted in universities and colleges, increasing the number of degree-trained chemical engineers many fold.

New chemical products and technologies transformed the postwar process industries. The principal products were petroleum and petrochemicals, synthetics, and pharmaceuticals.

During World War II, the refining of petroleum on a very large scale had developed rapidly. "Cracking" processes, producing shorter-chain derivatives from the long-chain molecules of heavy petroleum by thermal dissociation via a series of furnaces, yielded a variety of economically feasible products (see Cracking). The subsequent production of "petrochemicals," or chemical products synthesized from petroleum-based raw materials or feedstocks, developed most quickly in America, with its ample oil reserves. In petrochemical processes, petroleum is separated into various components by conventional distillation and catalytic cracking to yield a series of lighter compounds, which are separated, reacted with other compounds and combined to synthesize new chemical products.

By contrast, in Germany, which had been committed to a policy of self-sufficiency in resources during World War II, fuels had been based on coal rather than petroleum reserves. Britain, too, lacked established oil reserves in the first half of the century, and until the war its chemical industry relied on fermentation alcohol, coal, and tar derivatives for organic chemical production. Wartime demand and new products altered priorities. ICI, for example, previously a general chemicals manufacturer, had become a major producer of wartime aviation fuel in Britain. ICI's oil hydrogenation facility provided the basis of a petrochemical industry; Shell began producing a detergent based on wax cracking and after the war ICI sought associations with petroleum refiners to provide its feedstocks for nylon and polyethylene.

The development of new petrochemical products transformed the post-World War II chemical market, as much by replacement as by expansion. The business in solvents from sugarcane molasses,

for example, was undermined by the development of alcohol from ethylene. The alcohol market itself eroded in the early 1950s. Similarly, the dehydration of alcohol for use as motor fuel—an important process in the early years of the war—was abandoned. The coke oven industry withered with the employment of oil to replace coke in blast furnaces.

Like petrochemicals, the production of synthetic materials had prewar origins but came to have a dramatically greater economic influence after the war. Thermosetting products such as phenol-formaldehyde and epoxy resins, first produced in the 1890s, became practical products early in the twentieth century when their chemistry began to be studied. "Bakelite," for example, found a market from 1909 in the electrical and motor industries for insulating materials and compression mouldings. Similarly Rayon, based on cellulose, had been available from the 1890s and manufactured according to methods, and manpower, largely familiar to the traditional textile industry. Until World War I, most polymers were obtained from the chemical byproducts of gas works and coking ovens. Nylon, however, developed during the 1930s by DuPont in America, required production processes drawn from the chemicals industry. Its British counterpart, ICI, similarly established new processes for synthetic materials, from low-density polyethylene in 1939 (used extensively in wartime applications) to Terylene fibers (1949), Teflon (1950) and polyacrylonitrile fibers (1960). Here again, the developing discipline of chemical engineering promoted problem solving including heat transfer (because the reactions could proceed rapidly), filtration and mass transfer. The need for continued and intensive physical research was obvious: the characteristics of melt flow were distinct and complex for each new material, and the design of process equipment often a black art. Nevertheless, many such synthetic materials could be processed on equipment adapted from that developed earlier for the rubber industry for mixing, rolling and extruding.

By the mid-twentieth century there was a growing list of processes for producing chemicals using biological operations. Scientific brewing methods, for example, had flourished from the late nineteenth century. Several strands of research combined in the period before World War I to promote what has become known latterly as biochemical engineering or biotechnology. A rubber shortage between 1907 and 1910 encouraged searches for synthetic alternatives. Fermentation vied with distillation as a key unit operation in wartime production. After World War I, when petroleum reserves appeared limited, the production of synthetic fuels again helped to promote interest in biological methods. Nevertheless, the large manufacturers paid little attention to such processes in the interwar period, when the employment of coal-tar as a raw material appeared most promising.

World War II revitalized the development of biochemical processes. The production of synthetic rubber and pharmaceutical drugs assumed significant dimensions by the end of the war. The case of penicillin, in particular, illustrates how chemical engineering methods were adopted and adapted gradually, and how technology transfer operated between countries and disciplines. Penicillin production became the last major use of British agency factories during World War II. American pharmaceutical firms such as Pfizer, Merck, Squibb and Eli Lilly became major antibiotics manufacturers in the postwar years based on more efficient mould-growing technology developed there.

With the end of World War II, the international chemical industry was transformed. In 1951, the Allies broke up the IG Farben chemical cartel in Germany into a number of firms; the three largest—Bayer, Hoechst and BASF—independently extended manufacturing of pharmaceuticals, plastics and other chemical products. Similarly, the Swiss firm CIBA (Chemische Industrie im Basel), which had focused on textile dyes in the first half of the century, diversified into pharmaceuticals. The prewar activities of seeking oil and raw chemicals from colonies or possessions was extended from the 1950s to petroleum refining and chemical plants, notably in the near East and India.

By the 1950s, chemical engineers were beginning to provide a technical competence recognized in certain spheres, particularly in the design of plant for new processes. Indeed, the employment of chemical engineers in general plant contracting was an occupational niche that developed significantly after World War II. Contracting was an outgrowth of the independent consulting work and equipment brokering in which many early chemical engineers dabbled. The post-World War II version, however, was no longer small scale and short in duration. Nor was contracting any longer dominated by firms that also fabricated equipment. The American style of financing, design and fabrication of chemical plant was largely exported to Europe after the war.

The scale of production increased dramatically after the war, the capacity of new ethylene plants, for example, rising from some 10,000 tons per annum in 1950 to 450,000 tons per annum in 1970.

In Japan and West Germany, polymer production increased ten-fold during the 1960s. In the Soviet Union, conversion of the chemical industry from coal-based to petroleum-based feedstocks did not get underway until 1958. By 1970, some 90 percent of the world's chemicals were produced by the so-called "developed" countries.

Such expansion was based on continuous flow processes. On both economic and technical grounds, the refining of petroleum and the manufacture of petrochemicals were best undertaken as continuous processes. Instead of relying on the combination of premeasured components, flow processes required careful monitoring of reactants (which in turn depended increasingly on instrumentation such as flow meters and feedback control systems). Following the crude oil shortage of the early 1970s and a consequent quadrupling of prices, combined with growing public opposition in the West over pollution, such design elements allowed chemical process plants to be made increasingly efficient in the recovery and use of their byproducts.

The chemical industry had always been cyclic, and the demand for chemical products and industry workers by the early 1960s was made even more unstable by waves of investment. This economic cycle was exacerbated by delayed feedback: new investment in the increasingly large-scale chemical plants was followed by over-capacity, a reduction in profits and cuts in investment. Employment in the industry soared as plants were designed, constructed and started up, declined as a fewer operators were required to man them, and plummeted as further plant design was deferred. The rate of expansion of the international chemical industry slowed considerably by the 1970s. So too, did the number of jobs, journals, courses and professorial appointments, the rate of students graduating, and book publishing; all were to reach saturation by the early 1970s. During the last decades of the century this characteristic of restrained growth was sustained.

See also **Chemicals; Cracking; Detergents; Dyes; Feedstocks; Green Chemistry; Materials and Industrial Processes; Nitrogen Fixation; Oil from Coal Process; Plastics, Thermoplastics**

SEAN F. JOHNSTON

Further Reading

Aftalion, F. *A History of the International Chemical Industry*. University of Pennsylvania Press, Philadelphia, 1991.

Bradbury, F.R. and Dutton, B.G. *Chemical Industry: Social and Economic Aspects*. Butterworths, London, 1972.

Buchholz, K. Verfahrenstechnik (chemical engineering)—its development, present state and structures. *Soc. Stud. Sci.*, 9, 33–62, 1979.

Davis, G.E. *A Handbook of Chemical Engineering*, 2 vols. Davis Brothers, Manchester, 1901, 1904.

Divall, C. and Johnston, S.F. *Scaling Up: The Institution of Chemical Engineers and the Rise of a New Profession*. Kluwer, Dordrecht, 2000.

Haber, L.F. *The Chemical Industry 1900–1930: International Growth and Technological Change*. Clarendon Press, Oxford, 1971.

Hardie, D.W.F. and Pratt, J.D. *A History of the Modern British Chemical Industry*, Pergamon, Oxford, 1966.

Hayes, P. *Industry and Ideology: IG Farben in the Nazi Era*. Cambridge University Press, Cambridge, 1987.

Lesch, J.E., Ed. *The German Chemical Industry in the Twentieth Century*. Kluwer, Dordrecht, 2000.

Mowery, D.C. and Rosenberg, N. *Paths of Innovation: Technological Change in 20th-Century America*. Cambridge University Press, Cambridge, 1998.

Reader, W.J. *Imperial Chemical Industries: A History*, 2 vols. Oxford University Press, Oxford, 1970, 1975.

Reuben, B.G. and Burstall, M.L. *The Chemical Economy*. Longman, London, 1974.

Shreve, R.N. *Chemical Process Industries*. McGraw-Hill, New York, 1967.

Spitz, P.H. *Petrochemicals: The Rise of an Industry*. Wiley, New York, 1988.

Travis, A.S., Schröter, Harm G., and Homberg, E., Eds. *Determinants in the Evolution of the European Chemical Industry, 1900–1939: New Technologies, Political Frameworks, Markets, and Companies*. Kluwer, Dordrecht, 1998.

Chemicals

In the twentieth century, the chemical industry became an essential contributor to technological innovation, economic growth, and military power. At midcentury, *Fortune* magazine proclaimed that it was indeed a "chemical century." For the first half of the century, the industry had consisted of a diverse set of technologies serving a broad array of markets. The chemical industry provided thousands of products that were used in everything and everywhere. After the war the industry coalesced around products made from petroleum and natural gas, with plastics and polymers accounting for over half the industry's output. Another key growth area was pesticides, following on the example the wartime miracle chemical, DDT. The dramatic success of another wartime innovation, penicillin, led to the rapid growth of the modern pharmaceutical industry, based on a combination of chemistry and biology. In addition to specific products, the chemical industry provided scientific and technological knowledge needed to develop other critical

technologies, such as nuclear weapons and semiconductors. In the latter decades of the century, major innovations declined and competition increased causing chemicals to lose their high-tech image, reducing them to commodity status. The global industry has undergone massive reorganization in the wake of these new realities. Industry leaders hope that technologies based on biotechnology, green chemistry, and nanotechnology can restore the industry to its former glory.

Although people have made and used chemicals for thousands of years, the modern industry, based on large-scale production, emerged during the Industrial Revolution of the late eighteenth and early nineteenth centuries. The first industrial chemical—sulfuric acid—dates from the mid-eighteenth century when large lead-lined chambers were used to allow the oxidation of sulfur dioxide, made by burning sulfur, to sulfur trioxide, which reacts with water to produce acid. By the mid nineteenth century sulfuric acid plants had grown very large and had reached a high degree of technical sophistication, incorporating most of the techniques of modern chemical engineering. The availability of cheap sulfuric acid allowed the development of cheap alkali by the LeBlanc process, first developed in France but commercialized in Great Britain after 1810. Sulfuric acid was converted to sodium carbonate through a series of reactions with salt, limestone, and charcoal. Large quantities of acids and bases were consumed in Great Britain principally in textile operations, such as washing, bleaching, and dyeing.

Armed with these two reagents—acid and base—chemists began to experiment with a wide variety of substances, many of them organic (carbon-containing). By midcentury chemists had discovered some useful new compounds. In 1856, a young English chemist, William H. Perkin while naively trying to convert coal-tar into the valuable antimalarial quinine produced a purple colored solution instead. At the moment of this discovery extremely expensive purple was the fashionable color among Europe's elite. Using cheap coal-tar, a waste product from coal gasification plants that supplied illuminating gas to cities, Perkin developed a process to make a purple dye, mauve, by oxidizing aniline (benzene with an ammonia group substituted for a hydrogen atom). Other chemists soon discovered that the larger class of chemicals based on benzene rings would yield a rainbow of colors when reacted with acids and bases. The systematic and highly profitable exploitation of aniline dyes shifted in the 1870s to the new nation of Germany where the government, universities, and emerging chemical companies cooperated to develop this important industry. By World War I, three German companies, Bayer, BASF, and Hoechst controlled about 90 percent of the world's dyestuffs production. German chemists isolated the chemicals made by the madder and indigo plants that produced red and blue dyes, respectively. Chemists and engineers then learned how to manufacture these chemicals from coal-tar chemicals, replacing natural dyes which were major agricultural products of several countries, especially Turkey (madder) and India (indigo). Dyestuffs chemistry led German chemists into new fields such as pharmaceuticals with the discovery of aspirin by Felix Hoffmann and salvarsan (the first effective treatment for syphilis) by Paul Erlich. Another dyestuffs-related chemical, TNT (trinitrotoluene) would play a critical role as a shell-bursting explosive in World War I.

Explosives were revolutionized by chemists beginning in the middle of the nineteenth century. Experiments with nitric acid and organic molecules resulted in nitrate groups bonding onto the organic molecules, creating highly flammable or even explosive compounds. This characteristic resulted from the molecular proximity of a fuel (the organic compound) and oxygen (there are three oxygen atoms in each nitrate group). The most notorious of these new compounds was nitroglycerin, a liquid with tremendous explosive energy that was so unstable it often detonated prematurely. In Sweden, Alfred Nobel stabilized nitroglycerin by absorbing it into diatomaceous earth to produce a putty-like substance that could be extruded into paper casings. He called his product dynamite, and beginning in the 1870s it displaced black powder in blasting operations. Dynamite was one of the technological advances that would make projects such as the Panama Canal feasible.

Even more important than dynamite was its chemical cousin, nitrocellulose, prepared by reacting nitric acid with cotton fibers. This still cotton-like material became the basis for smokeless powder, which in the 1890s began to replace black powder as the propellant in guns and cannon. Smokeless powder burned much more cleanly than black powder and was much more powerful. The new propellant made the machine guns and heavy artillery into the terribly effective weapons that turned World War I into a bloody stalemate. Smokeless powder had a tendency to decompose causing spontaneous fires and explosions, until German chemists discovered a dyestuffs-related compound that stabilized the powder in 1908. Another key chemical in the munitions

machine was TNT, which exploded on shell impact causing huge craters and saturating the air with shrapnel.

The ingenuity of chemists added another horrific element to life in the trenches—poison gas. At the Battle of Ypres in 1915, German chemist Fritz Haber orchestrated the release of 5000 cylinders of chlorine which drifted with the wind into the Allied lines. The burning, choking gas caused panic in the Allied army but the Germans were not prepared to attack and so lost the advantage of its new weapon. Afterward both sides used poison gases such as lewisite, phosgene and mustard gas throughout the remainder of the war. All of these gases contained chlorine, which could be made in large quantities using electrochemical technology.

The development of the dynamo in the 1870s made available large quantities of electricity that could be used to make chemicals, many of which could not be economically made by other methods. Perhaps, the most important example was aluminum, which had semiprecious metal status—a small pyramid of it capped the Washington Monument which was completed in 1883. Three years later, Charles Martin Hall in the U. S. and Paul Louis Toussaint Heroult in France discovered a process to make aluminum using electricity. This method is still used today.

Another important electrochemical process was the production of chlorine and caustic soda (sodium hydroxide) from salt water. Chorine was used principally in bleaching powder and sodium hydroxide became the major base, replacing earlier compounds such as sodium carbonate. This process began to be used in 1890s; several electrolytic plants were built near Niagara Falls where hydroelectric power was available, and Herbert Dow built an early plant in Midland, Michigan where there was a rich supply of brine wells.

Other important materials were made in electric furnaces, which could generate very high temperatures, invented by Henry Moissan in 1892. One new ceramic compound was silicon carbide, which is so hard that it can be used to shape metals by grinding. Another was calcium carbide which reacts with water to produce acetylene, used in early automobile head lights and in oxyacetylene metal-cutting torches. Made from coal, acetylene became an early chemical building block used to make other chemicals.

The development of the Haber–Bosch ammonia process between 1906 and 1912 was a technological and scientific tour de force that became a prototype for future chemical processes. One of the great scientific and technological challenges of the late nineteenth century was "fixing nitrogen." Nitrogen was an essential ingredient in explosives and fertilizer. Most of the world's useable nitrogen came from nitrate mines in the Atacama desert in northern Chile. Of course, air is 80 percent nitrogen, but it is almost chemically inert because it consists of two tightly bound atoms. Chemists sought ways to break those bonds. One way to do this was to react nitrogen and hydrogen to make ammonia. On paper it looked simple; in the laboratory it did not happen under normal conditions. A solution to this apparent impasse was suggested by theoretical considerations derived from the evolving disciplines of kinetics (the rate of chemical reactions) and chemical thermodynamics (determines the feasibility of particular reactions). The ammonia reaction was found to be feasible by German chemists Walter Nernst and Fritz Haber. Their calculations showed that the reaction would occur at very high temperatures (for kinetics) and very high pressures (for thermodynamics). The challenge then became technological: was it possible to build steel vessels that could withstand temperatures of 500°C and a pressure of 200 atmospheres? After Haber was able to make ammonia in laboratory scale apparatus, Carl Bosch of the BASF Company oversaw the development of a commercial process. Some of the early reactors were made from Krupp cannons. An essential part of the process was the development of a catalyst, a substance that causes the nitrogen and hydrogen to react with each other. At BASF, Alwin Mittasch led an exhaustive search until an efficient and durable iron-based catalyst was developed. The first large plant started up in 1913 a year before World War I would make Chilean nitrates unobtainable in Germany because of Britain's dominance of the seas. Without "synthetic" nitrogen, the Germans could not have sustained their war effort for four years.

In the 1920s BASF would expand on its high-temperature, high-pressure technological base by developing processes to make methanol from carbon monoxide and hydrogen and gasoline from coal. Before the new process, methanol was obtained by distilling it from wood (hence its name wood alcohol). The synthetic gasoline project was initiated by predictions of impending shortages of crude oil. After 1929, the discovery of the east Texas oil field increases world crude supplies and the Great Depression lowered demand for gasoline, the huge investment in synthetic gasoline technology threatened the viability of the giant IG Farben chemical combine. (The major German chemical companies had merged in 1925 primarily

to sustain export markets.) The project and company would be rescued by Hitler after he came to power in 1933, since a domestic supply of gasoline—Germany has no oil—would be essential in a future war.

Hitler's policy of autarky sustained another project that would have important consequences for the chemical industry—synthetic rubber. Making synthetic versions of natural materials had been a long-standing objective of the chemists and one of the foundations of the chemical industry. Dyestuffs had been the first major success, but chemists also sought to make other substances, especially silk and rubber. Until the 1920s the basic structure of these substances was a matter of scientific uncertainty. This, however, did not stop chemists from forging ahead trying to make synthetic substitutes for exotic and expensive natural materials.

The origin of synthetic materials dates to 1870 when Albany tinkerer, John Wesley Hyatt formed a solid plastic from a mixture of nitrocellulose and camphor, which he called celluloid. According to tradition, Hyatt was looking for a substitute for expensive elephant ivory in billiard balls. When his new material failed in this use, he then made celluloid look like exotic materials—ivory, amber, and tortoiseshell—so it would be used in toilet sets, toys, and numerous other trinket-like applications. Its most enduring legacy was as the film base that made motion pictures possible beginning in the 1890s. An unsuccessful use of nitrocellulose was as an artificial silk fiber that, among other deficiencies, was highly flammable.

A much better silk-like fiber was rayon, formed by dissolving cellulose to make a syrupy viscose solution that was extruded through small holes in a plate into another chemical bath that solidified the fiber. Charles Cross and Edward Bevan in Britain discovered this process in the 1890s, while attempting to make improved light bulb filaments. After 1910 the market for rayon fibers began to expand rapidly worldwide; the fashion industry embraced it the 1920s; and during the Great Depression it replaced silk in all apparel except stockings. Rayon was the biggest new product for the chemical industry in the interwar years.

Rayon was just one a growing number of products made of large molecules (or macromolecules), in this case it was natural cellulose. Chemists were beginning to make entirely new large molecules. A pioneer is this effort was Leo Baekeland who invented a hard plastic he dubbed Bakelite in 1907. The new material was made by heating phenol and formaldehyde under pressure. Among the many uses for Bakelite was as a substitute for ivory in billiard balls.

The growing importance of and interest in large molecules in the 1920s sparked a scientific debate, especially in Germany—still the center of chemistry—about their structure. Although many chemists argued that large molecules were held together by peculiar forces, Hermann Staudinger put forth the hypothesis that large organic molecules were just that—larger versions of common organic chemicals held together by same types of chemical bonds. Following Staudinger, Wallace H. Carothers, a researcher in the DuPont Company developed methods for making large molecules, or polymers, in the laboratory. Out of this research DuPont researchers discovered neoprene synthetic rubber (1930) and nylon (1934). By 1940 neoprene had established itself as a specialty rubber and nylon had become the preferred stocking fiber. Once the mysteries surrounding polymers had been solved, chemists everywhere began to explore this large and promising new field.

Perhaps the most significant discovery, both historically and for the future chemical industry, was made in 1929 by IG Farben chemists who made a general purpose synthetic rubber from a polymer consisting of repeating units of butadiene and styrene. At the time of this breakthrough virtually all of the world's rubber came from British controlled plantations in Malaysia. By early 1942, these were all in Japanese occupied territory. The first year of American fighting was hampered by a lack of rubber which threatened to bring the effort to a thudding halt. To resolve this crisis the U.S. government organized a cooperative venture between oil, chemical, and tire companies to rapidly build up an American synthetic rubber capability. This initiative was a marked success, production went from nothing to 800,000 tons in two years. One of the major obstacles that had to be overcome was to develop processes to make enormous quantities of styrene and butadiene. Styrene was available before the war in limited quantities but butadiene was not a commercial chemical. The supply of butadiene came primarily from oil companies, which had previously concentrated on making fuels not chemicals.

In the interwar years a few companies such as Union Carbide and Shell Oil had begun to make chemicals from petroleum and natural gas. One notable product introduced in the 1930s was ethylene glycol—automobile radiator coolant antifreeze. Until World War II organic chemicals used as feedstocks for the chemical industry were distilled from coal. For example, the type of

nylon DuPont commercialized was determined mainly by the abundance of benzene, a major coal impurity. After World War II the oil and chemical industries, especially in the U.S., would soon shift to petrochemical feedstocks.

The oil industry was now generating large quantities of chemicals as a byproduct of new processes developed to produce more gasoline from a barrel of crude of oil and to produce higher octane fuels, especially necessary for aviation gasoline during World War II. Crude oil is a complex mixture of hydrocarbons with varying carbon atom chain lengths. Originally, natural gasoline, which contains five to nine carbons in each molecule was distilled out of crude oil. In 1913, E. M. Burton developed a cracking process in which he subjected the heavier fractions of crude oil to heat and pressure which broke the larger molecules into smaller ones, some of which were in the gasoline-size range. In the 1930s, French inventor and engineer, Eugene Houdry, added a catalyst to this process which significantly improved its overall performance. A decade later, an improved catalytic process called fluidized bed cracking was developed by Massachusetts Institute of Technology chemical engineers and Standard Oil of New Jersey. This process has been used ever since. Also during the late 1930s oil companies began to develop processes to combine some of the smaller cracked molecules into larger ones that could boost the octane rating of gasoline. During World War II, American 100-octane aviation fuel helped Allied pilots win the air war over Europe.

A few years after the war ended, the Universal Oil Products company, a research organization, introduced a new process which had been developed by Vladimir Haensel. Called "platforming" because of its platinum catalyst, it dramatically improved the octane rating of gasoline primarily by stripping hydrogen atoms from cyclical compounds, converting them to benzene, tolulene, and xylene. These compounds were not only important for high-octane gasoline but were in great demand by the chemical industry as raw materials, especially for the booming plastics and polymers businesses.

The dramatic post-World War II expansion of the chemical industry was led by plastics and polymers. Shortages of metals and other materials during the war had prompted the U.S. government to encourage manufacturers to use plastics for a wide variety of applications. For example vinyl resin (mostly polyvinyl chloride or PVC) production increased from 2.3 million to 100 million kilograms. Although many plastics ended up in applications such as army bugles, others served essential high-technology functions. Polyethylene, a difficult to make plastic, had unique insulating properties needed for radar. DuPont's exotic polymer Teflon, which did not melt, dissolve in solvents, or stick to anything, was used as a sealant in the Manhattan Project for the atomic bomb. Clear acrylic plastic became the material for airplane windows and bomber gunner turrets.

After the war both the uses and varieties of polymers increased to fulfill the demand by consumers for whom convenience became a hallmark of modern life. A New England inventor, Earl Tupper, introduced his line of polyethylene food storage containers—Tupperware—that preserved leftovers and kept them neatly in the refrigerator. The most sensational new products were synthetic textile fibers that made clothing more affordable, machine washable, drip dry, and wrinkle free. DuPont's nylon took over the stocking market and made major inroads in other apparel. Polyester, discovered by two British chemists in 1940, was used instead of wool in suits and blended with cotton to make permanent press garments. Acrylic fibers made sweaters, especially lightweight ones, popular with postwar women. By 1956, synthetic fibers had eclipsed wool as the number two textile fiber consumed in the U.S. By 1970, synthetic fiber consumption surpassed that of cotton in apparel.

At the same time that synthetic fibers were revolutionizing modern wardrobes, new types of plastics found myriad uses. The major breakthrough in plastics was made when German chemist, Karl Ziegler in 1953 discovered a new type of catalyst that produced new kinds of polymers, notably linear polyethylene and polypropylene. These two plastics had outstanding properties such as toughness which led to many uses, especially food packaging. As polymer science matured in the 1950s, chemists made more complex and sophisticated compounds, examples being DuPont's Kevlar polyaramid and Lycra spandex fibers.

Another significant growth sector for the post-World War II chemical industry was organic-chemical based pesticides. The archetype was DDT, whose remarkable kill-on-contact property was discovered by Paul Mueller in Switzerland in 1939. Most earlier insecticides were poisons, such as lead arsenate, that had to be ingested. During the war, George W. Merck, working on biological warfare for the government, discovered that a DuPont plant growth compound called 2,4-D was actually an effective herbicide. After the war chemical companies focused research efforts on finding new insecticides, herbicides, and fungicides.

Although Rachel Carson's *Silent Spring* (1962) publicized the toxic effects of DDT on birds and raised questions about the effect of pesticides on human health, during that decade 96 new insecticides, 110 new herbicides, and 50 new fungicides were introduced. Insecticides included organophosphorus compounds, carbamates, and synthetic pyrethrins. DuPont in 1967 introduced Benlate (benomyl), the first fungicide that was taken up internally rather than being effective only on the leaf surfaces. In the 1970s Monsanto introduced its blockbuster herbicide, Roundup (glyphosate), which was suitable for a wide variety of crops. In the 1990s chemical companies, especially Monsanto and DuPont combined biotechnology and herbicides to create crop seeds that were compatible with specific herbicides, and to incorporate insecticidal properties into plants by splicing in genes from other organisms. These so-called genetically modified foods have created controversy in Europe but have met with little resistance in the U.S.

It became evident in the 1960s that the chemical industry was maturing. During the 1970s the industry was beleaguered by spikes in the cost of energy and feedstocks, and environmental legislation that required major capital investments in pollution control and abatement. By the 1980s, the chemical industry had become very competitive worldwide with growth and profits tightly linked to larger business cycles. Since then the industry had undergone massive reorganization in response to these new economic realities. For the most part, chemicals, if not the companies that make them, have become commodities. Because it still has significant research capabilities, the chemical industry is hoping that new technologies such as nanotechnology—very small molecular structures—or green chemistry—replacing petroleum with renewable feedstocks—might restore chemicals to the essential status it enjoyed in the twentieth century.

See also **Chemical Process Engineering; Dyes; Electrochemistry; Explosives, Commercial; Feedstocks; Green Chemistry; Nanotechnology; Nitrogen Fixation; Oil from Coal Process; Pesticides; Synthetic Resins; Synthetic Rubber; Warfare, Chemical**

JOHN KENLY SMITH

Further Reading

Arora, A., Landau, R. and Rosenberg, N., Eds. *Chemicals and Long-Term Economic Growth: Insights from the Chemical Industry*. Wiley, New York, 1998.

Haber, L. *The Chemical Industry, 1900–1930: International Growth and Technological Change*. Clarendon Press, Oxford, 1971.

Haynes, W. *American Chemical Industry: A History*. Van Nostrand, New York, 1945–1954.

Hounshell, D.A. and Smith, J.K. Jr. *Science and Corporate Strategy: DuPont R&D, 1902-1980*. Cambridge University Press, Cambridge, 1988.

Idhe, A.J. *The Development of Modern Chemistry*. Harper & Row, New York, 1964.

Spitz, P.H. *Petrochemicals: The Rise of an Industry*. Wiley, New York, 1988.

Chromatography

The natural world is one of complex mixtures, often with up to a 100,000 (e.g., proteins in the human body) or 1,000,000 (e.g., petroleum) components. Separation methods necessary to cope with these, and with the simpler but still challenging mixtures encountered, for example in pharmaceutical analysis, are based on chromatography and electrophoresis, and underpin research, development, and quality control in numerous industries and in environmental, food, and forensic analysis.

In chromatography, a sample is dissolved in a mobile phase (initially this was a liquid), which is then passed through a stationary phase (which is either a liquid or a solid) held in a small diameter tube—the "column." According to the differing relative solubilities in the two phases, mixture components travel through the column at different rates and become separated before emerging and detected by the measurement of some chemical or physical property. The sample size can be as small as one picogram (10^{-12} g), but tens of grams can be handled in preparative separations.

Chromatography spans the twentieth century: the separations described by the Russian botanist Mikhail Tswett in 1903 by continuous adsorption/desorption in open packed columns, commonly applied in natural-product chemistry, was followed by the Nobel Prize-winning work (awarded 1952) of Archer John Porter Martin and Richard Laurence Millington Synge. Tswett called his new technique "chromatography" because the result of the analysis was "written in color" along the length of the adsorbent column (in this case, chalk). In 1941 Martin and Synge replaced liquid–liquid extraction by liquid partition chromatography (LC), supporting the liquid stationary phase on a solid support (using silica gel), over which they passed the sample solution. In their paper they made the famous statement:

> "The mobile phase need not be a liquid but may be a vapor... very refined separations of volatile substances

> should therefore be possible in a column in which permanent gas is made to flow over gel impregnated with a nonvolatile solvent."

This forecast led to the reality of gas chromatography (GC), developed by Martin and his co-worker Anthony T. James in 1952. Their first GC comprised of a glass tube containing the support material through which a gas mixture was blown by a carrier gas. The separated analytes in the mixture were detected using a simple titration technique.

LC is applicable to all soluble substances, whereas GC can be applied to mixture components which can be heated without decomposition to give sufficient vapor pressure (measured as a few millimeters of mercury, mmHg) for each solute to pass through the column in reasonable time. The volatility range of GC can be extended by the use of a supercritical fluid (a substance above its critical temperature and pressure) as mobile phase—supercritical fluid chromatography (SFC). This is a better solvent than a gas and more diffusive than a liquid, hence permitting more rapid analysis. Capillary electrochromatography (CEC) combines the advantages of LC and electrophoresis. The mobile phase flow is maintained by an electric field rather than by applied pressure; the separation principle is again partition, but the effect of different rates of electromigration may be superposed in applications to charged analytes.

Early partition chromatography was carried out in columns with an inert support on which the stationary phase was either coated or, better, bonded. In 1958 M.J.E. Golay showed how a tortuous path through a packed bed could be replaced by a much straighter path through a narrow open tube. Long, and hence highly efficient columns could thus be fabricated with the stationary phase on the inner wall, and remarkable separations achieved by GC and later by SFC. LC took a different course because slow diffusion in liquids means that separations in open tubes require impractically small diameters. Increased efficiencies can be more simply achieved on columns packed with small silica particles with bonded organic groups.

Initial commercial GC instrumentation was based on glass packed columns, and the most famous gas chromatograph was the Pye 104 instrument with a flame ionization detector. The direct coupling of gas chromatography and time-of-flight (TOF) mass spectrometry was achieved in the mid-1950s in collaboration with W.C. Wiley, I.H. McLaren, and Dan Harrington at the Bendix Corporation. At about the same time, GC was coupled to a magnetic sector instrument. The great utility of modern GC-MS was made possible by the advent in the 1960s of carrier gas separators that removed the GC carrier gas prior to introduction of a sample into the high-vacuum mass spectrometer. In 1979, the landmark paper by Ray Dandeneau of Hewlett Packard announced the development of fused silica columns to the world. Capillary columns have more theoretical plates (a measure of column resolving power or efficiency) per meter as compared to packed columns and since they have less resistance to flow they can be longer than packed columns. This means that the average capillary column of 30 meters has approximately 100,000 theoretical plates while the average packed column of 2 meters has only 2500 plates.

High-pressure liquid chromatography (HPLC) was developed in the mid-1970s and quickly improved with the development of spherical silica packing materials and the small volume UV detector. In the late 1970s, reverse phase liquid chromatography gave improved resolution and the technique was widely accepted into the pharmaceutical industry.

By the 1980s HPLC was commonly used for the separation of small molecules. New instrumentation techniques improved separation, identification, and quantification far above the previous techniques. Computers and automation added to the convenience of HPLC. Since 1990 there has been a continuous movement from the 4.6-millimeter internal diameter columns of the 1970s and 1980s to microbore columns ranging in internal diameter from between 1 and 3 millimeters. Today this is moving into capillary columns that range from 3 to 200 microns. These capillary columns have provided an easier interface to the mass spectrometer that is now very quickly becoming the detector of choice for LC.

Other forms of chromatography are:

- Thin-layer and paper chromatography. First developed in the early work by Martin and Synge, where a layer of absorbent is spread on a glass plate, and mixtures of components are placed on the edge of the plate. Solvent is then allowed to move up the plate by capillary action, drawing the components of the mixture along by varying degrees.
- Gel-permeation chromatography. Compounds are separated on the basis of their molecular size.
- Ion-exchange chromatography. Separation of ions based on their charge using a charged support material.

Electrochromatography, capillary electrophoresis (CE and CZE), capillary gel electrophoresis (CGE), micellar electrokinetic capillary chromatography (MECC) and capillary electro chromatography (CEC) are new separation techniques employing high voltages and narrow bore capillaries they are all grouped into the area of electrochromatography. This technique utilizes the principles and advantages of electro-osmotic flow.

See also **Electrophoresis**

KEITH D. BARTLE AND PETER MYERS

Further Reading

Giddings, J.C. *Dynamics of Chromatography Principles and Theory*. Marcel Dekker, New York, 1965.

Cazes J. and Scott, R.P.W. *Chromatography Theory*. Marcel Dekker, New York, 2002.

Grob R.L. and E.R. Barry, Eds. *Modern Practice of Gas Chromatography*. Wiley, New York, 2004.

Neue, U. *HPLC Columns*.Wiley, New York, 1997.

Lee, E.M.L. and Markides K.E., Eds. *Analytical Supercritical Fluid Chromatography and Extraction*. Chromatography Conferences, Provo, Utah, 1990.

Leslie, S.E. *Milestones in the Evolution of Chromatography*. ChromSource, Franklin, Tenn, 2002.

Bartle, K.D. and Myers, P., Eds. *Capillary Electrochromatography*. Royal Society of Chemistry, Cambridge, 2001.

Brinkman, U.A.T., Ed. *Hyphenation: Hype and Fascination*. Elsevier Science, Amsterdam, 1999. Reprinted from: *J. Chromatogr*. A, 856, 1–2, 1999.

Mondello, L., Lewis, A.C. and Bartle K.D., Eds. *Multidimensional Chromatography*. Wiley, New York, 2001.

Civil Aircraft, Jet Driven

The development of jet power shortly before World War II resulted in a series of transport projects designed during the conflict, especially in England. The jet principle seemed simple in theory. Air drawn through an intake is compressed through the engine nacelle by turbine-powered rotating blades, mixed with injected kerosene, which is then burned to create energy and rear-flowing exhaust gas. In fact, the internal engine dynamics, such as compression rates, temperature, and the fuel injection system all required considerable experimentation, with military jets becoming the first to be powered that way.

In the U.K., the Brabazon Committee, a government air transport task force formed in 1942, determined in the course of its second gathering a year later that several new types of airliners would be needed after the war. Among these, "type IV" was to be a jet-driven plane, a challenge the De Havilland Company picked up with its model 106, soon known as the Comet. The machine in question went through several design changes, yet while news of its construction for British Overseas Airways (BOAC) was common knowledge, no pictures or drawings were shown to the media or the public until the machine's ground trial runs. First flown in 1949, the Comet was not the only civilian machine of its kind (Avro Canada flew its C-102 that same year), nor the first (a Vickers Viking test aircraft was), but it became the most visible as it outpaced all other projects under development, entering service in 1952.

The Comet began to revolutionize air travel by radically cutting time spent in the air, while offering a new level of comfort associated with higher altitude cruising. However, two accidents in 1954 grounded all Comet aircraft. When the cause of the crashes was discovered (unforeseen metal fatigue), the redesign of the machine delayed the introduction of the transatlantic version of the plane by four years, thus giving the advantage to a Boeing model that would set new standards for jet travel: the 707.

Initially derived from a military air refueling project, the 707 also signaled new trends in commercial aviation. First, it signaled the beginning of the dominance of the jet market by American manufacturers, one which would not be successfully challenged until the Airbus consortium marketed several new aircraft in the 1980s. Second, it reflected a shift in airliner manufacturing practices, whereby constructors learned to offer multiple versions of the same basic model to increasingly selective airlines who might turn to the competition otherwise. Boeing, for example, had to redesign the initial 707 with a wider fuselage after United Airlines chose the Douglas DC-8 competitor aircraft. Boeing also worked on a slightly smaller version of the plane, known as the 720, to meet airline demand for a faster plane over a shorter range. Airlines eager to draw on an elite clientele quickly relegated their propeller-driven planes and introduced jets which, thanks to their speed, became new heralds of modernity. New models quickly followed.

While jet plane design was focused first on placing the engine within or attached to the wing, other solutions were also devised beginning in the 1950s to deal with the matter of thrust, noise, and aerodynamics. In the realm of short- and medium-range aircraft (1500 to 4000 kilometers), the French-built Caravelle became the first aircraft to use rear-mounted twin engines. A solution that reduced noise considerably in the passenger cabin

and simplified the design of the wing, the model was highly successful for its time, and inspired such a design on several other models, such as the Douglas DC-9. Eventually, the latter together with the Boeing 737 twin (with engines under the wings) took back much of the Caravelle market, thanks in part to the greater flexibility of the American manufacturers in designing fuselage extensions in response to airline demand. However, other manufacturers, like the Dutch Fokker company and the British Aircraft Corporation maintained a European presence with the Fokker 28 and the BAC 1-11 short- to medium-range machines. These, together with updated Boeing and Douglas aircraft formed the second generation of jet airliners. By then, airports had lengthened their runways to cope with jets' higher take-off speed and greater payload, and new designs, such as mobile gangways leading directly from the waiting area to the plane were being designed.

This second generation, appearing in the late 1960s and early 1970s, also became associated with mass travel due to the introduction of wide-bodied aircraft. The development of a new type of jet engine, the turbofan, improved the thrust ratio 20 to 1, with the new engines amounting to only 4 percent of the total aircraft weight (against 9 percent for the 707). This massive increase in thrust made the completion of the such planes as the three-engined Douglas DC-10 and Lockheed Tristar, and the giant Boeing 747 possible.

The 747 was originally designed in response to a U.S. military call for a heavy (over 350 tons) long-distance jet transport. Boeing lost the bid to Lockheed's C-5A Galaxy plane, but was able, thanks to Pan American Airways' initial order, to develop a commercial wide body (characterized by twin aisles with up to ten seats per row). Able to carry some 350 passengers across the Atlantic in one fell swoop, later versions of the 747 would sport a capacity increase up to 450 passengers (with a consequent reduction in comfort and seating space). Boeing also offered two Japanese carriers a version designed for frequent take-offs and landings intended to link Japanese cities commuter-style, which could seat 550 people. Generally, however, the 747 and other wide-bodied long-range aircraft became associated with affordable travel beginning in the late 1970s, when airline deregulation allowed for greater competition.

However, early in its career the 747 was often found to be much too big for many links. The airline business thrives on filling seats, and it was not uncommon that the 747 flew well under capacity. While the DC-10 and L-1011 remedied this problem (they sat on average 250 passengers at capacity) the added troubles of worldwide recession and the double oil shocks of the 1970s meant that airlines began requesting smaller machines that would also use less fuel. The twin-engine jets that appeared in the early 1980s filled that gap in all ranges.

The first aircraft manufacturer to recognize this niche was the European Airbus consortium, which began to offer the A-300 with up to 250 seats in 1972. Although sales were close to nil at first (analysts suggest the machine came on the market too soon), by the late 1970s a smaller, more versatile model for up to 200 passengers, the A-310, began picking up sales. This third generation jetliner faced as main competition Boeing's new 767 of comparable size. More recent versions of both planes, were used to lengthen the ETOPS time rule (Extended Twin Engine Operations) that required aircraft with fewer than three engines to always be within 120 minutes of an airfield in case of single engine failure. With the advent of other twinjets, both short range and long range, the struggle between Airbus and Boeing has continued to this day, and spread into the realm of fourth-generation airliners.

The latest generation of jetliners, though outwardly similar to earlier ones, is characterized by "glass cockpits" (where liquid crystal displays have replaced dials and buttons), fly-by-wire systems that initially existed on military aircraft alone (computers send electronic signals to the controls), and more efficient engines with more complex wings (to minimize fuel consumption and noise pollution). Such technological advances have contributed to reducing the number of cockpit occupants from four for the first generation jetliners, to three by the second (the navigator disappeared), to two by the latest (the engineer is no longer needed). At the same time, however, the increase in distances covered on some flights (16,000 kilometer stretches are now routine), have made the inclusion of a relief crew essential. At the same time, the increase in passenger numbers, the need for revenue, and the competitive market have replaced the veneer of prestige characterizing early jet travel with visions of delays, poor service, and air traffic congestion.

See also **Aircraft Design; Turbines: Gas, in Aircraft**

GUILLAUME DE SYON

Further Reading

Davies, R.E.G. *A History of the World's Airlines*. Oxford University Press, Oxford, 1964.

Davies, R.E.G. and Birtles, P. *De Havilland Comet*. Paladwr Press, McLean, 1999.
Dierikx, M. and Bouwens, B. *Building Castles of the Air: Schipol Amsterdam and the Development of Airport Infrastructure in Europe, 1916–1996*. Sdu Publishers, The Hague, 1997.
Gunston, B. *Airbus: The European Triumph*. Osprey, London, 1988.
Heppenheimer, T.A. *Turbulent Skies: The History of Commercial Aviation*. John Wiley, New York, 1995.
Jarrett, P. Ed. *Modern Air Transport. Worldwide Air Transport from 1945 to the Present*. Putnam, London, 2000.
Newhouse, J. *The Sporty Game*. Alfred A. Knopf, New York, 1982.
Sabbagh, K. *21st-Century Jet: The Making and Marketing of the Boeing 777*. Scribner, New York, 1996.

Civil Aircraft, Propeller Driven

The advent of early commercial propeller aircraft followed in the wake of World War I, when manufacturers modified bomber models in response to airline demand. Thus, the Farman Company developed the Goliath model, which was used in early Paris to London links beginning in 1919. Along similar lines, using a modified Vickers Vimy, British pilots Alcock and Brown were able to cross the Atlantic that same year. However, their exploits and those of Charles Lindbergh notwithstanding, propeller aviation in the interwar years focused primarily on short and medium range aircraft. As engine and airplane design improved the range of machines, some were used for long-distance mail transport.

With every nation, including defeated Germany, eager to invest in civil aeronautics, hundreds of civil aircraft initially derived from military machines appeared in European skies. The most famous machine built specifically for transport purposes, however, was the German Junkers F-13. As the first all-metal airliner it proved extraordinarily sturdy and, combined with the Junkers Company's heavy investment in foreign airlines, helped spread the use of that machine. Its small size, however (it carried only four passengers), meant that newer models would be required as traffic demand expanded. Other manufacturers such as Fokker supplied very successful machines, including the three-engined F-VII. Other manufacturers offered single-engine machines, like the Latécoère 28 or the Potez 25, both in service on French airmail routes.

In the interwar years, the push to expand aircraft range came primarily from lucrative airmail contracts. In both Europe and the U.S., new solutions were thus devised to solve the dual problem of speed and distance covered. In North America, a mix of indigenous and foreign machines shared the scene, while in Europe, long distance fascination was focused on the Atlantic Ocean.

Early airlinks with South America were run partly by boat, but modified planes like the Latécoère 28 "Comte de la Vaulx" successfully bridged the South Atlantic in 1930, though Germany sought to check French expansion by building an airship service. In addition, the Dornier company tried (but failed) to develop the Giant Dornier DO-X to fly across the Atlantic regularly (it once lifted 169 passengers on a demonstration flight), but had greater success supplying the German Lufthansa with Dornier Wal ("Whale") aircraft, which were refueled after a sea landing by specially positioned ships in the middle of the South Atlantic. Other midrange solutions included the British experimentation with a tandem system involving the Short S-20 "Mercury" being launched from a "Maia" mother plane. With the exception of the German system, most such solutions remained experimental.

By the 1930s, despite European progress in the realm of aircraft such as the Dewoitine 338, the Handley Page Hermes, the Junkers 52 Trimotor and the Focke-Wulf 200 Condor, the primary thrust towards the modern airliner was to be found in the U.S. There, the Ford Trimotor, derived from a freight model, was used for coast-to-coast passenger service in two days beginning in 1929. Four years later, however, the Boeing Company designed and flew what is considered the first modern airliner. Its model 247 was a twin-engine all-metal machine that could carry ten passengers, and flew on average 120 kilometers per hour faster than other airplanes in service, thanks to its improved aerodynamics. Airline and aircraft manufacturing competition, however, led the Douglas Company to take the next step when it agreed to design a twin-engine plane, the DC-1, for TWA. Improvements to this model led to the introduction of the DC-2, which had a considerable range for a serial-built land plane at the time (some 1000 kilometers). Other manufacturers entered the fray, yet in the realm of twin-engine transport Douglas remained an uncontested leader with the introduction of its next model, the DC-3, which could accommodate 21 passengers. Used heavily in World War II, a few surviving examples of the more than 10,500 built remain in service to this day.

In other categories, several manufacturers tried their hand at speed, with Lockheed offering the Orion model, and Heinkel building the He-70. In such cases, however, the design of the machines

sacrificed comfort, and proved more useful for mail transport than passenger service.

In parallel to the development of land planes, giant flying boats were studied and tested over long distances. Deemed safer than land planes and faster than airships, these appeared to offer the ideal solution to the problem of long-distance passenger transport. England, France, Germany, and the U.S each witnessed the development of several types of flying boats after the failed DO-X, some with six engines, but World War II interrupted most testing and the projects became military ones. A notable exception was the Boeing 314 model, in service with both Pan American Airways and BOAC, which allowed for a wartime service between the U.K. and the U.S.

During World War II, the main developmental thrust came from the U.S., where the war effort sped up already existing projects. Douglas, which had flown its four-engine DC-4 model in 1938, completed development of the machine, and Lockheed began development of its Constellation model. Consequently after World War II, European airlines eager to catch up bought American planes, despite pressure from their respective governments to support national industries. By 1947, pressurized versions of the Constellation as well as an improved version of the DC-4, the DC-6, and the introduction of the Boeing 377 Stratocruiser signaled further dominance of the long-range market by U.S. machines (the pressurized cabins allowed flights above 3000 meters, thus avoiding most of the dangerous weather systems).

In the medium- and short-range markets, France and Great Britain were able to produce a few piston-engined propeller aircraft, but none was as successful as the DC-3, which remained in use throughout the 1950s and early 1960s.

The propeller era would have likely come to an end with the advent of the jet were it not for the development of the turboprop engine, in which instead of gasoline-powered engines with heavy cylinders and pistons, new, kerosene-fuelled turbines moved the propeller. The result was that while fuel consumption was on average slightly higher than a piston engine, the energy output could be almost double for the same weight (thus increasing speed and payload), and the safety level far greater. Several turboprop airliners were thus developed, including the British Vickers Viscount, the Soviet Tupolev 114, and the American Lockheed Electra. While longer-range turboprops gave way to jets, the short- and medium-range market held their ground, and new models appeared in the 1960s, including the Fokker 27 and the British Aerospace HS-748. In Canada, the De Havilland Twin Otter became a favored machine for its rugged performance and capacity to operate from short runways in cold weather, and gained worldwide success.

Finally, in the 1980s, several models were developed to service the commuter market, including the Swedish-built Saab 340 twin, and its main competitor, the Italo–French ATR 42. Although later developments in the small jet plane market may displace such turboprop planes, their capacity to operate from short runways will keep such machines in the market for years to come, especially in remote areas where airport development is more challenging.

See also **Aircraft Design; Civil Aircraft, Jet Driven; Internal Combustion Piston Engine; Turbines, Gas**

GUILLAUME DE SYON

Further Reading

Davies, R.E.G., *A History of the World's Airlines*. Oxford University Press, Oxford, 1964.

Dierikx, M. *Fokker, a Transatlantic Biography*. Smithsonian Institution Press, Washington D.C., 1997.

Gwynn-Jones, T. *Farther and Faster: Aviation's Adventuring Years, 1909–1939*. Smithsonian Institution Press, Washington D.C., 1988.

Heppenheimer, T.A. *Turbulent Skies: The History of Commercial Aviation*. Wiley, New York, 1995.

Jarrett, P., Ed. *Biplane to Monoplane. Aircraft Development 1919–39*. Putnam, London, 1997.

Jarrett, P. Ed. *Modern Air Transport. Worldwide Air Transport from 1945 to the Present*. Putnam, London, 2000.

Propliner Magazine quarterly, 1979–present.

Stroud, J. European post-war piston engined airliners. *Aeroplane Monthly*, a monthly series, 1992–1995.

van der Linden, F.R. *The Boeing 247: The First Modern Airliner*. University of Washington Press, Seattle, WA, 1991.

van der Linden, F. R. *Airlines and Airmail: The Post Office and the Birth of the Commercial Aviation Industry*. The University Press of Kentucky, Lexington, 2002.

Civil Aircraft, Supersonic

When the jet age began revolutionizing air travel by halving flight time and doubling passenger numbers in the 1950s, supersonic transport (SST) appeared to be the logical next stage in the progress of commercial aviation But while the dynamics and challenges of supersonic flight were clearly defined (the speed of sound, Mach 1, ranges between 1000 and 1200 kilometers per hour)), applying them to civilian transportation remained extremely difficult. The challenges included propulsion, weight

distribution (on military aircraft the payload is easily predetermined), passenger comfort and, as it would soon become clear, economic efficiency. Several nations undertook studies of supersonic transport in the late 1950s including the U.S., the Soviet Union, the U.K, and France.

In the U.S., an SST project was announced by President Kennedy in 1963 and saw the involvement of major aerospace giants Boeing and Lockheed. However, while Boeing gained the upper hand, the proposals it offered for the design of a 300-seat aircraft proved too unrealistic. Public and political impatience, combined with skyrocketing costs, prompted a termination of governmental funding in 1971 and the project was abandoned, having reached a total cost of US$1 billion.

In the Soviet Union, too, several designs for SSTs were considered, with Tupolev receiving the go-ahead to design the model 144. First flown on 31 December 1968, ahead of the Anglo-French Concorde, the machine was also the first passenger airliner to exceed Mach 2, and was shown at several Paris Air Shows. However, the machine required considerable redesign between the first and the second prototypes, and several problems, such as cabin noise, were never satisfactorily solved. One prototype crashed at the 1973 Paris Air Show, but testing continued and service with the Soviet airline Aeroflot did happen, albeit briefly, in the late 1970s between Moscow and Alma Ata. The production aircraft were then withdrawn from service and stored, though one saw testing service through a National Aeronautics and Space Administration (NASA) grant in the 1990s.

In the U.K, the impulse towards SST studies began in the mid-1950s, when the Ministry of Supply formed a Supersonic Transport Aircraft Committee. The STAC suggested two types of SST projects, one medium range, one long range, with speeds ranging from Mach 1.3 to 2.0. By 1959, the long-range option was favored as a means to make British aerospace industry more competitive in the jet age, since Boeing had taken the lead over the ill-fated British Comet jetliner. British company Bristol had gained considerable experience in the jet engine realm, however, by developing the Olympus military motor. The staggering cost projection for such an aircraft, however, meant that when the British government awarded the British Aircraft Corporation (BAC) a development contract in 1960, it recommended that international partners be found to help cushion the investment risk. Openings to American companies revealed a different design philosophy, which projected either a Mach 3 airliner, or a 300-passenger giant using expensive titanium. Across the Channel, however, French state manufacturer Sud Aviation, based in Toulouse, had done studies parallel to those of BAC. Following heavy technical and political negotiations, a deal was signed in 1962 for the joint development of an SST, whereby the engine development was 60:40 in favor of Britain, while the airframe work proportion was reversed in favor of France.

While more financial and political hurdles lay ahead (the project was almost cancelled in 1965 by the British Labour government of Harold Wilson), technical challenges became some of the toughest any aircraft constructor had to overcome. Among these were the extreme temperatures a Mach 2 aircraft would experience while cruising at an altitude high above conventional jetliners (18,000 meters instead of 9,000). This kinetic heating would extend the fuselage some 16 centimeters despite the intense cold, and different sections would heat up to temperatures between 90 and 120°C. The windows consequently had to be designed extremely small to avoid stress problems.

Similarly, the British Olympus engines had to function at a variety of speeds, and provide acceleration with afterburners. They were based on military engines, but required further refinement to carry passengers and control fuel consumption. For example, the air rushing into the engines at Mach 2 would actually be too fast and its shock wave would destroy the powerplant. Engineers came up with an ingenious system of variable size inlets and doors that changed the boundary layer and slowed the air down during cruising speed (Figure 2). The gas dynamics of the

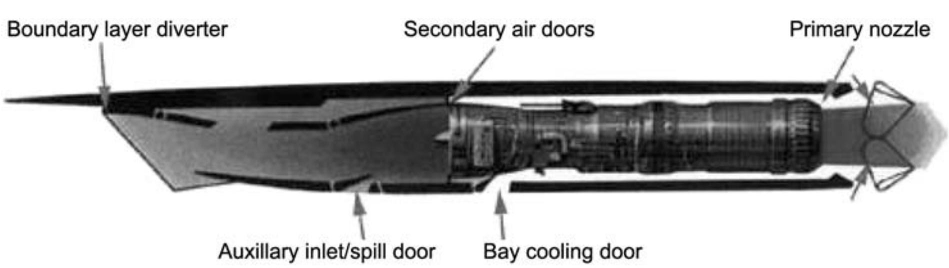


Figure 2. Supersonic transport engine.

air also meant that at higher altitudes, drag would be reduced, and the engines would have to do less work. However, reaching the desired altitude involved considerable thrust, and therefore dealing with higher exhaust temperatures. On subsonic aircraft, the exhaust gas necessary to provide the thrust comes out through a narrowing nozzle. In the case of Concorde's Olympus engine, the converging nozzle was assisted by a diverging, or expanding nozzle that provided the extra thrust necessary for the plane to carry a payload.

The question of gas exhaust remained central to the development of the engines. If a Concorde cruises supersonically at about 700 meters per second, then the exhaust gases must leave the engine at least at 900 meters per second. During acceleration and climb, this could be done through afterburning, a process common on military jets, where fuel is ignited into the exhaust gases, producing the necessary energy. This would become the favored solution for take-off and acceleration. It also meant higher temperatures would affect the engines, ranging from 120°C at the inlet level, to 550°C in the compressor area, followed by even higher temperatures when fuel was burned. This required the use of titanium and nickel alloy for the compressor blades. In the case of Concorde, most of the plane is made of Hiduminium RR58, an aluminum alloy, but some areas, including the engine exhausts and areas surrounding them were made of high-strength steel and titanium; these metals would have been the primary choices had Concorde been designed to reach Mach 3, but their weight made them less likely to be selected.

The aircraft's needle shape impaired cockpit vision considerably, especially on the ground. To allow the crew to maintain visual contact with the runway during take-off and landing, a hinged nose and visor were designed to drop as low as 12 degrees in relation to the cockpit windshield, and fully retract during supersonic cruise. Other innovations included the transfer of fuel during different flight phases from one tank to another. As the plane emptied its tanks, its center of gravity would shift, which would affect its speed. Thus, in addition to the main fuel tanks situated in the wings, forward and rear trim tanks were added as a means to control the center of gravity.

The Concorde project called for the testing of two prototypes and two pre-series aircraft (designed to iron out any last minute details in assembly and flight operations, and serve as airline demonstrators). The projected production was at least 150 machines, and when the prototypes flew in 1969, sixteen major world airlines had taken options on over 70 aircraft. Yet by the time manufacturing closed down in 1979, some 16 production Concordes had been built and delivered to the only two airlines that bought it, British Airways and Air France. The causes for the commercial failure are multiple, but can be traced back to the uneconomical features of the machine, its pollution level, and the oil shocks of the 1970s. As one journalist once remarked, when compared to an early Boeing 747, which first flew the same year as Concorde did, the supersonic flew twice as fast, but consumed four times more fuel to carry four times fewer passengers (about 100). Second, as aircraft testing was underway in the 1970s, new environmental concerns began to appear in the industrialized world, including matters of atmospheric pollution and airport noise abatement. Although Concorde was not significantly noisier than other planes of its time while taxiing at an airport, its noise print at cruising altitude raised such concerns that ultimately it would only be allowed to fly at its maximum speed over oceans. Finally, the oil shocks of 1973 and 1979 provided the proverbial nails in the coffin by making fuel extremely expensive and accentuating a world economic downturn.

At the level of operations, both British Airways and Air France faced a difficult start that involved wrestling American landing rights for the machine, trouble phasing the supersonic into their respective fleets, and selecting which destinations would become money makers. Operations began in 1976, and by the mid-1980s, both airlines were making a profit on the Concorde, though this eventually required limiting their service to scheduled Paris and London to New York and Washington schedules, and luxury charter flights (Figure 3).

The crash of an Air France Concorde in July 2000 appeared to toll the end of the first passenger supersonic era, but following safety modifications to the fuel tanks, the flying certificate was reissued, and both airlines resumed operations. In 2003, however, the airline industry's economic downturn prompted both operators to announce the end of operations that year. As for a "son of Concorde," multiple projects have been announced for years, but thus far they have consistently encountered problems similar to the supersonic pioneer's, including cost and concern for the environment.

See also **Aircraft Design; Civil Aircraft, Jet Driven**

GUILLAUME DE SYON

Figure 3. Concorde flying over Manhattan (New York City).
[*Courtesy of Adrian Meredith at www.concordephotos.com.*]

Further Reading

Davies, R.E.G. *Supersonic (Airliner) Non-Sense: A Case Study in Applied Market Research*. Paladwr Press, McLean, 1998.

Horwitch, M. *Clipped Wings: The American SST Conflict*. MIT Press, Cambridge, MA, 1982.

Knight, G. *Concorde, The Inside Story*. Stein & Day, New York, 1976.

Moon, H. *Soviet SST: The Technopolitics of the Tupolev-144*. Orion Books, New York, 1989.

Trubshaw, B. *Concorde: The Inside Story*. Sutton Publishing, Phoenix Hill, 2000.

Cleaning: Chemicals and Vacuum Cleaners

Chemicals

During the nineteenth century, rising standards of cleanliness, a new breed of domestic engineers, an increase in the number of objects people owned, and the emergence of the ''germ theory'' influenced the ideology and practice of cleaning. Still, throughout the twentieth century, people continued to clean with many of the same chemicals as had their predecessors in earlier centuries. The distribution of piped hot water and electrical current contributed as much to a shift in cleaning practice as any change in the nature of chemicals. Twenty-two percent of British households lacked piped hot water as late as 1961. The impact of soaps and cleaning agents that require hot water for effective use must not be overestimated.

Before the twentieth century, many cleaning agents were made at home. While manufactured pastes, liquids and polishes were available by the early nineteenth century to those who could afford them, both the novelty and the effectiveness of these have been questioned. Soda played a significant role in housecleaning in the twentieth century, in part because it was cheaply and easily acquired after Frenchman Nicholas Leblanc developed an efficient way of mass producing it from salt in the late eighteenth century. Historian Susan Strasser claims that only two cleaning agents were manufactured before 1880. Besides soda, twentieth century women continued to rely on sand, milk, salt, borax, camphor, lye, vinegar, turpentine, clay, acids, and oils.

The spread of dishwashing and laundry chemicals followed the introduction and mass manufacture of the dishwasher and washing machine, most prominently in Europe and North America after World War II. These products were often simply refinements of existing ones. Liquid soap was introduced by Minnetonka in the 1980s. In the latter three decades of the twentieth century, manufacturers responded to consumer concerns about the effect of their products on the natural environment. Chlorine, petrochemicals and phosphates were among the agents avoided in the new environmentally friendly cleaning products.

Vacuum Cleaners

The vacuum cleaner was a product of a desire for greater cleanliness and arguably one of the technologies that most abetted greater cleanliness. Along with the electric iron, the vacuum became one of most widely owned household appliances in the twentieth century. Historians agree, however, that the dissemination of the vacuum barely affected time spent in cleaning. Yearly or bi-yearly outdoor carpet and curtain shakings were replaced by the daily or weekly use of the vacuum, as standards of cleanliness rose. The history of the vacuum cleaner remains ''an important example of the commercial application of the phobia against dirt'' (Forty, 1982).

The portable electric vacuum did not appear for sale until the first decade of the twentieth century. By the middle of the nineteenth century, portable appliances that sucked particles by means of

bellows and others with both air draft and revolving brushes had been patented. Those who could afford to, sent their carpets out to be professionally cleaned. Steam-powered industrial machines simply beat the carpet with rubber beaters. Later, machines employed steam to destroy insects and a rotary fan to blast dust out of the carpet and up a chimney. An American machine that blew compressed air through carpets inspired British civil engineer William Booths, who patented a "suction machine" in 1903. Booth's "Puffing Billy," which could be installed permanently in a building or mounted on a trolley and pulled by horses and automobiles, was exceptional among dozens of similar inventions in that it was the only one with a self-contained power source.

American David T. Kenney had developed a similar machine in the U.S., and its installation as a central vacuum in the Frick house in Manhattan set a precedent which was followed in both domestic and commercial applications. Central vacuums would not achieve widespread use, however, because their capital and installation costs were prohibitive. In addition, most American or European households had not been electrified at the time of its invention. Chapman and Skinner in San Francisco invented the first "portable" electric vacuum in 1905. It weighed 42 kilograms and used a fan 45 centimeters in diameter to produce the suction. Because of its size, it did not sell well. By 1908 consumers could purchase the heavy (18 kilograms) but portable Hoover Model O Electric Suction Sweeper, for $75.00, a considerable sum at the time. It consisted of a tin body, filled with a fan, a motor and a rotating cylindrical brush. Behind the motor, a bag attached to the handle received the refuse. This model included a flexible nozzle that could be fitted to the machine and used to vacuum upholstery. By 1926, the motor and parts were much lighter and a beating mechanism had been added to assist the brushes and suction. By this time, the machine was known simply as the "Hoover." A second type of portable electric unit made by the Swedish Electrolux Company eliminated the rotating brushes from the design and used only suction. The Model O and the Electrolux provided the functional and design paradigms for vacuum cleaners until the end of the twentieth century.

Improvements and refinements through the end of the twentieth century included more efficient and quieter motors; lighter (and later recyclable) plastic parts; disposable dust bags; flexible cord winders (electrical cords that could be "sucked" back into a neat bundle); microfilters; and streamlined designs. In 1979, Black and Decker Company introduced the cordless vacuum, the battery-operated and soon very popular "Dustbuster." British designer and engineer James Dyson developed a bagless vacuum, which sucked and filtered dust into a plastic container that replaced the bags that Dyson claimed became clogged with dust and prevented efficient absorption. After marketing what became the fastest-selling vacuum in Britain, Dyson began working on a robotic vacuum.

See also **Detergents; Dishwashers; Laundry Machines and Chemicals**

JENNIFER COUSINEAU

Further Reading

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.

Davidson, C. *A Woman's Work is Never Done: A History of Housework in the British Isles 1650–1950*. Chatto & Windus, London, 1982.

Forty, A. *Objects of Desire: Design and Society from Wedgwood to IBM*. Pantheon Books, New York, 1986.

Gideon, S. *Mechanization Takes Command: a Contribution to Anonymous History*. W.W. Norton, London, 1948.

Hardyment, C. *From Mangle to Microwave: The Mechanization of Household Work*. Polity Press, Oxford, 1988.

Lupton, E and Miller, J.A. *The Bathroom The Kitchen and the Aesthetics of Waste: A Process of Elimination*. Princeton Architectural Press, New York, 1992.

Lupton, E and Miller, J.A. *Mechanical Brides: Women and Machines From Home to Office*. Princeton Architectural Press, Princeton, New Jersey, 1993.

Nye, D. *Electrifying America: Social Meanings of A New Technology*. MIT Press, Cambridge, MA, 1990.

Strasser, S. *Never Done: A History of American Housework*. Pantheon Books, New York, 1982.

Yarwood, D. *Five Hundred Years of Technology in the Home*. B.T. Batsford, London, 1983.

Clocks and Watches, Quartz

The wristwatch and the domestic clock were completely reinvented with all-new electronic components beginning about 1960. In the new electronic timepieces, a tiny sliver of vibrating quartz in an electrical circuit provides the time base and replaces the traditional mechanical oscillator, the swinging pendulum in the clock or the balance wheel in the watch. Instead of an unwinding spring or a falling weight, batteries power these quartz clocks and watches, and integrated circuits substitute for intricate mechanical gear trains.

When quartz timepieces first hit the market, it seemed unlikely that the expensive gadgets would sell. Instead they won over consumers and revolu-

tionized the way timepieces are made, sold, and used. Today, quartz is the most common source of time and frequency signals, not only in clocks and watches but also in scientific instruments and in other consumer products like computers, cell phones, and television sets.

The technology that made quartz a practical time base for domestic clocks and watches developed from several independent research streams that stretched over nearly a century. Quartz, the mineral silicon dioxide, is one of Earth's most common materials. It has special properties, as the Curie brothers, Pierre and Jacques, had shown in the late nineteenth century. When subjected to electrical voltage, quartz vibrates, many thousands of times a second at a regular rate that is dependent on how it is cut and shaped (this phenomenon is known as piezoelectricity). During the 1920s, amateur radio operators, many of whom made their own crystal units, demonstrated the potential of quartz for controlling broadcast frequencies. At the same time, industrial scientists were also investigating the properties of quartz for frequency control.

The first quartz clock, constructed in 1927, was nearly as big as a room. Canadian telecommunications engineer Warren Marrison and his colleague J. W. Horton at Bell Telephone Laboratories in New York City developed the basic system that subsequent quartz timepieces would employ. To keep time, the clock counted the vibrations of a quartz crystal in an electrical circuit, subdivided those vibrations—the equivalent of a mechanical clock's "beats"—to minutes, seconds and hours that were displayed on a dial. In building his clock, Marrison had been searching for ways to monitor and maintain precise electromagnetic wave frequencies that carry radio and telephone messages. When he found that the crystal's vibrations were stable enough to hold electromagnetic waves to a particular frequency, he also realized the potential of his invention for improved timekeeping.

After Marrison, others built quartz clocks and demonstrated repeatedly that they were more accurate than the best mechanical clocks. During the 1930s and 1940s, influential research institutions throughout the world—including the Naval Observatory and the National Bureau of Standards in the U. S., the Royal Greenwich Observatory and the British Post Office in the U. K., and the Physikalisch-Technische Reichsanstalt in Germany—installed quartz clocks. These clocks remained rare, expensive and experimental scientific instruments through the 1940s, when scientists began developing atomic standards even more accurate than quartz. At the same time, commercial quartz standards, built by such firms as General Radio Company of Cambridge, Massachusetts, served the needs of those with slightly less precise time and frequency needs than the world's elite laboratories—mainly broadcast facilities and electrical power stations.

World War II laid the groundwork for the postwar quartz revolution in the global watch industry. Because of its neutral status, Switzerland continued to make timepieces while countries engaged in the war converted their watch factories to supply military customers. As a result, Switzerland responded to high civilian demand, exported cheap mechanical movements throughout Europe and the U.S., and gained a seemingly indomitable lead in the global watch industry. American watch companies—like Hamilton, Elgin, and Bulova—survived the war years through military contracts or by assembling cheap Swiss movements into American-made cases. Meanwhile, watch production in Japan nearly ceased altogether. After the war, only a few watch firms remained worldwide, and those seeking to challenge the Swiss for a share of the market began to investigate alternative technologies potentially more accurate than the best mechanical watches.

World War II anticipated the postwar quartz revolution in another way. The enormous demand for radio communications during the war stimulated both the production of quartz oscillator units and research into the material's characteristics and behavior. This established quartz as a viable source of time and frequency signals. But quartz units and wartime electronics technology based on vacuum tubes were both still too large and power-hungry for watches and domestic clocks. The postwar advent of transistors and integrated circuits ultimately provided components small enough to be alternatives to gears, springs and tubes.

The history of the modern electronic wristwatch began in the 1960s when, in pursuit of more accurate timepieces, teams of engineers—working independently in Japan, Switzerland and the U.S.—used newly created microelectronic components to completely reinvent the wristwatch. The three teams that ultimately brought the first quartz watches to market took three completely different approaches to the task.

At K. Hattori & Company (now Seiko Corporation) in Japan, a group formed in 1959 began by building a quartz clock, and through successive products, miniaturized their quartz timekeepers to the size of a chronometer, a pocket watch and ultimately a wristwatch. Their Seiko

Astron SQ, introduced in Tokyo on Christmas Day 1969, was the first quartz watch sold anywhere in the world.

The Swiss were only a few months behind. In 1962 the industry had founded a new research laboratory, Centre Electronique Horloger (CEH) in Neuchâtel, to develop a new kind of electronic watch. By 1967 researchers there had two kinds of working quartz watch prototypes, which, along with Seiko's, made their debut at time trials at the Neuchâtel Observatory in 1967. The first Swiss quartz watches for sale—each of which contained an electronic module designed at CEH and dubbed Beta 21—were available on April 10, 1970 under the brand names of nearly 20 different watch companies.

The Americans were even further behind. The firm responsible for the first quartz watch produced in the U.S., the Hamilton Watch Company of Lancaster, Pennsylvania, launched its quartz watch project in 1967. Because the Swiss and Japanese had a clear lead in quartz analog watches, those with the traditional dial and hands, Hamilton concentrated its efforts on inventing the first solid-state digital watch. Partnering with Electro-Data, a Texas electronics company, Hamilton brought out the Pulsar in 1972. At the push of a button, the watch displayed the time of day with flashing red numerals made from light-emitting diodes (LEDs). It cost $2100—the price of a small car.

Domestic quartz clocks appeared on the market about the same time as the Pulsar, and, even though they were larger than watches, they too depended on the miniaturization of electronics. The earliest examples came from the German firms Junghans, Kienzle, and Staiger, the British firm F. W. Elliot and, in the U.S., General Time's Westclox brand. Already by the mid-1970s, quartz timepieces began to shift from expensive rarities to ubiquitous low-cost timepieces.

The introduction of the first quartz watches disrupted the entire global watch industry. Established watch firms were slow to respond, and new U.S. semiconductor companies—Texas Instruments, Fairchild, and National Semiconductor, for example—began making huge numbers of digital quartz watches, with both LED and liquid crystal displays (LCDs). Competition quickly forced prices down. In Japan, Seiko, Citizen and Casio quickly invested in new electronic watch equipment, and for a time they led in worldwide watch production. By 1978 Hong Kong, concentrating on low-end modules that cost pennies to make, exported the largest number of electronic watches worldwide. Both traditional watch manufacturers and the U.S. semiconductor companies competing in the consumer electronics market were irreparably damaged. The Swiss watch industry suffered near-fatal decline, reorganized completely, and became competitive again in the 1980s, buoyed by an entirely new product—the Swatch.

Consumers also played a role in this quartz revolution. The earliest quartz watches had, along with high prices, significant shortcomings. Watch buyers complained about short battery life, large clunky cases, and technical breakdowns caused by hard knocks and moisture. LED watches were difficult to read in daylight, and LCDs could not be seen at night. Interactions with users over these issues forced manufacturers to improve their products.

Consumers weighed in on the way the new watches displayed time too. The electronic wristwatch evolved almost overnight from a rare gadget for rich men to a cheap throwaway timepiece for men, women, and even children. Easy availability coupled with the introduction of digital watch displays, radically different in appearance from the traditional dial and revolving hands, generated passionate responses from consumers, both positive and negative. Some serious opponents of digital displays in the 1970s even predicted the day would come when the traditional dial would vanish completely and our ways of knowing time would be irretrievably diminished. The debate over the relative suitability of digital versus analog displays continues to this day. Since the mid-1980s, sales figures have shown most buyers prefer the analog dial over the digital display.

By the 1990s nearly all new watches and clocks made were electronic. As a result, nearly everyone has access, whether we need it or not, to the split-second accuracy once available to only scientists and technicians.

See also **Integrated Circuits; Transistors**

CARLENE E. STEPHENS AND MAGGIE DENNIS

Further Reading

Glasmeier, A.K. *Manufacturing Time: Global Competition in the Watch Industry, 1795–2000*. The Guilford Press, New York, 2000.

Johnstone, B. *We Were Burning: Japanese Entrepreneurs and the Forging of the Electronic Age*. Basic Books, Boulder, 1999.

Landes, D. *Revolution in Time: Clocks and the Making of the Modern World.* The Belknap Press of Harvard University Press, Cambridge, MA, 2000.

Marrison, W. The evolution of the quartz crystal clock. *Bell Sys. Tech. J.*, 27, 510–588, 1948.
Stephens, C. and Dennis, M. Engineering time: inventing the electronic wristwatch. *Br. J. Hist. Sci.*, 33, 477–497, 2000.
Weaver, J.D. *Electrical and Electronic Clocks and Watches.* Butterworth, London, 1982.

Clocks, Atomic

Atomic clocks rely on the precise timing of the oscillation frequency of certain atoms which are energized or excited to another energy state. The near immutability of the energy levels in the atomic structure of certain paramagnetic elements provides atomic clocks with the inherent reproducibility and long-term stability previously lacking in material standards such as quartz clocks or the Earth's rotation period. Some ten thousand (commercial) atomic clocks are now in operation throughout the world ensuring essential control and synchronization in a wide range of applications in science and technology, such as in the NAVSTAR satellites used in the global positioning system (GPS). The readings of a subset of more than 250 clocks maintained in about 60 national institutions are combined with those of about 12 laboratory clocks to form International Atomic Time (TAI), the basic time scale for science and technology and whose derivative, Coordinated Universal Time (UTC), which differs from TAI by an integral number of seconds, provides the atomic equivalent of mean solar time for general use worldwide.

The principle underlying atomic clocks is that when individual atoms of the same element absorb or emit energy by electrons moving from one energy level to another, the radiation produced by individual atoms has exactly the same (quantized) frequency. These resonant frequencies, stable in time and space, are which makes the atoms perfect timepieces. The oscillations are changed by environmental factors such as humidity that would decrease the accuracy of normal clocks.

The main sources of atomic timekeeping are cesium and rubidium clocks with a smaller, but significant, contribution from active hydrogen maser clocks. In the atoms of all three elements the ground state (lowest-energy state) is split into two hyperfine levels by the interaction between the spin (and consequent magnetic moment) of the single valence electron and the spin of the atomic nucleus. The energy difference, $\triangle E$, between these two closely separated states determines a resonant or characteristic frequency ν given by $\nu = \triangle E/h$, where h is Planck's constant. Transition from the top to the lower state causes emission of a photon of that frequency. The states are nearly equally populated and the necessary unbalance; that is, alteration of the distribution of atoms in a given state, can be achieved by either magnetic state selection involving spatial separation of the atoms (selecting only those in one hyperfine state) or optical pumping using differential transfer to excited states.

The magnetic approach was first employed in the 1920s when Otto Stern and Walther Gerlach in Germany made use of high-gradient magnetic fields to deviate a beam of (silver) atoms, each hyperfine state being oppositely deflected, confirming space quantitization. The method was subsequently improved in the late 1930s by Isodor Rabi at Columbia University who added a radio-frequency field to stimulate transitions between the hyperfine states and thus brought together the elements of a magnetic resonance machine that could function as a frequency standard. Those atoms that are stimulated to change state emit light. The frequency of the radio-frequency field is varied, and eventually, a frequency is achieved that alters the states of most of the atoms and maximizes their fluorescence. A counter counts the radio-transmitter pulses, which gives a precise frequency standard that can be used to define a time standard.

In 1949 Rabi's colleague Polykarp Kusch produced a basic design concept for a cesium atomic clock, his proposals incorporating a Ramsey cavity, following the discovery in the same year by Norman Ramsey at Harvard University that coherent excitation applied at the beginning and end of an interval of beam travel was equivalent to atom-field interaction over the whole interval, thereby reducing substantially the width of the resonance line.

In the following years work began at several laboratories on cesium beam resonators based on the Kurch design. In the U.S. Harold Lyons at the (then) National Bureau of Standards achieved a linewidth of 300 hertz in 1952 with Ramsey excitation. At the Massachusetts Institute of Technology (MIT) Jerrold Zacharias developed a compact, transportable clock in 1954 which would form the basis for the first commercial production in late 1956. Meanwhile, at the U.K. National Physical Laboratory (NPL), Louis Essen and John Parry, largely following the Kurch recipe, brought a cesium beam standard into operation in June 1955. This date represents the start of atomic timekeeping for, unlike the U.S. developments the NPL resonator was linked to an existing quartz

clock ensemble and the resulting atomic clock could be routinely compared with astronomical time. The calibration of the cesium frequency in terms of Ephemeris Time over a three-year period led in 1967 to the formal redefinition of the second as 9,192,631,770 periods of the cesium-133 hyperfine transition frequency.

The present uncertainty in the realization of the second by laboratory standards making use of magnetic selection is about part in 10^{14}, equivalent to nearly 1 nanosecond (10^{-9} seconds) per day while the commercial clock contributing to TAI display stabilities of the same order. The essential features of a commercial resonator are shown in Figure 4. Atoms traveling at about 200 meters per second are selected by magnet A in one of the hyperfine levels and make the transition ($F = 4$, $m = 0$)($F = 3$, $m = 0$), the Zeeman substates $m = 0$ having least dependence on the low magnetic field in the C region. Thereafter, the atoms are directed by magnet B to a surface ionization detector, evaporating as positive ions which provide the servo-output to drive the frequency of the exciting oscillator to the peak of the Ramsey pattern.

Alfred Kastler in 1950 proposed a scheme of optical pumping which would be fully exploited only many years later with the advent of diode lasers. However, the partial overlap in the spectra of the rubidium isotopes ^{85}Rb and ^{85}Rb enabled optically pumped rubidium gas cell frequency standards and clocks to be realized in the late 1950s. The main features are shown in Figure 5. The two D lines at wavelengths of 780 and 795 nanometers in the ^{87}Rb lamp in (a) are differentially absorbed in the ^{85}Rb filter cell (b), the residual light allowing transitions only from the F=1 hyperfine level (c) in the vapor cell, thus depopulating the level and making the cell more transparent. Microwave excitation at the hyperfine frequency equalizes the hyperfine populations again, resulting in increased light absorption and a consequent fall in detector output, thereby providing a control signal to the microwave source. A buffer gas in the cell shield the rubidium atoms from depolarizing contact with the cell walls and largely eliminated Doppler broadening of the resonance line.

In the 1950s Zacharias at the Massachusetts Institute of Technology had constructed a vertical cesium resonator with the intention of achieving a long Ramsey interval using slow atoms interrogated as they rose and fell under gravity through a single cavity. He was not successful using thermal atoms but the “Zacharias fountain” was finally realized by André Clairon and his colleagues at the Paris Observatory in 1993, making use of atoms cooled to a few microkelvin, state-selected by optical pumping and then periodically projected upwards at speeds of only a few meters per second to give a resonance width of about 1 hertz. Several fountain clocks are now in operation and

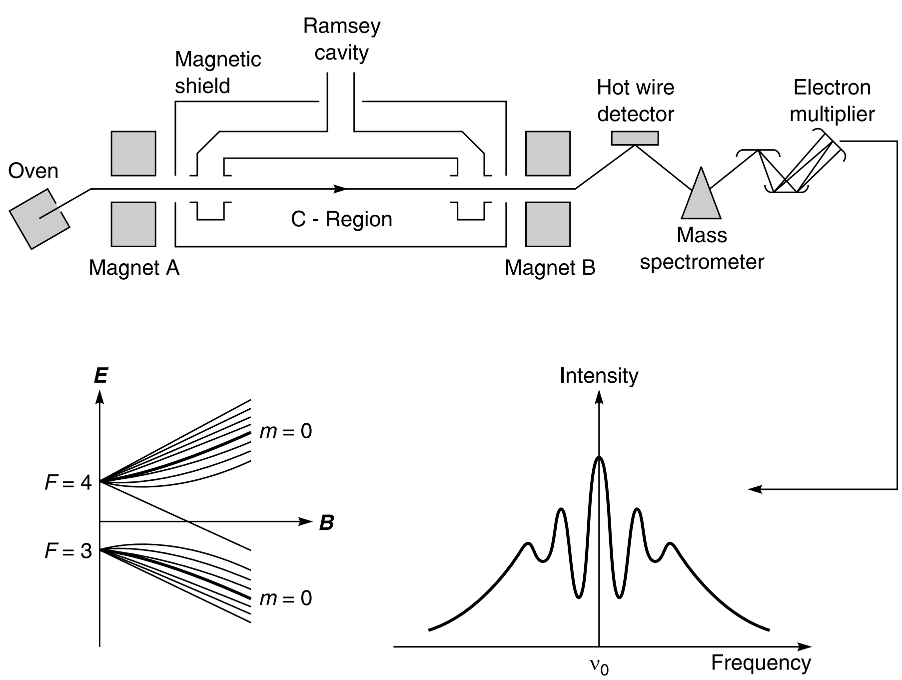


Figure 4. Features of a commercial resonator. [*Source: Audoin, C. and Vanier, J. Atomic frequency standards and clocks, J. Phy. E: Sci. Instrum., 9, 697–720, 1976.*]

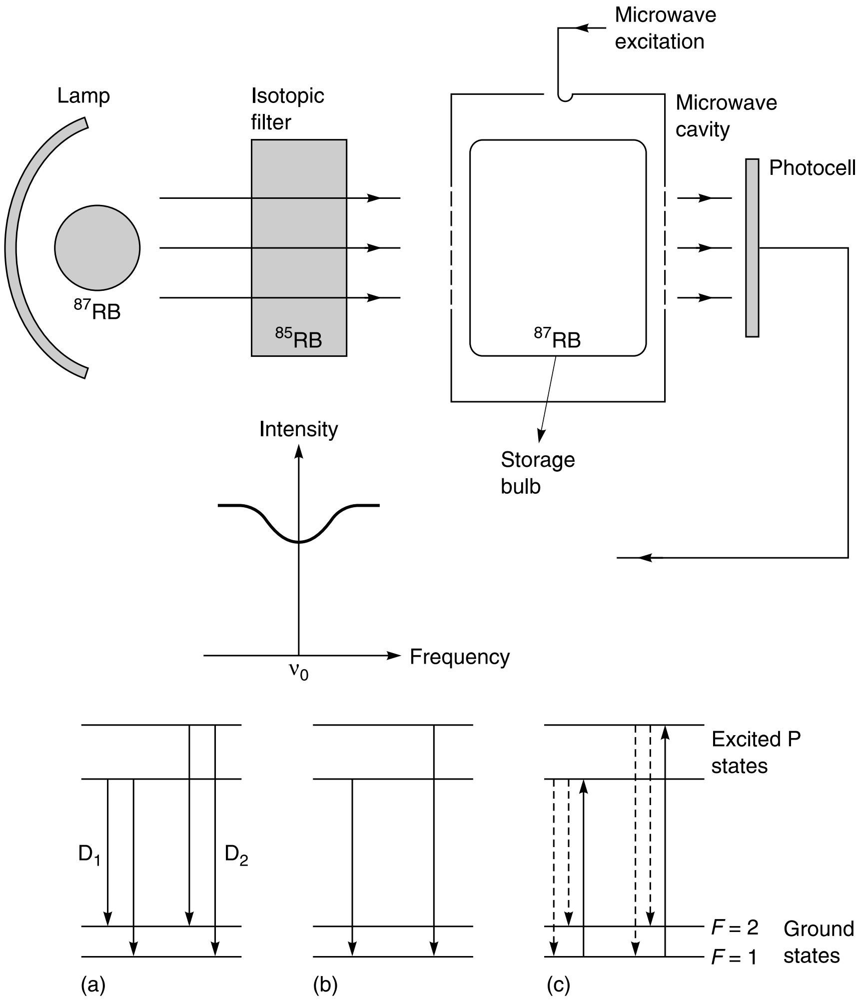


Figure 5. Features of optically pumped rubidium gas cell frequency clocks.
[*Source: Audoin, C. and Vanier, J. Atomic frequency standards and clocks, J. Phy. E: Sci. Instrum., 9, 697–720, 1976.*]

contributing to TAI with an uncertainty of about 1 part in 10^{15} equivalent to 100 picoseconds (10^{-12} seconds) per day.

The clocks so far considered have been passive devices requiring external excitation. The atomic hydrogen maser, shown in outline in Figure 6, first operated in Ramsey's laboratory at Harvard in 1960. It produces a signal at about 1420 megahertz albeit at the low level of around 10^{-12} watt.

Magnetic state selection focuses atoms in the $F = 1$, $m = 0.1$ states into the bulb immersed in a low-loss cavity resonator. A film of Teflon applied to the wall of the bulb allows atoms to make thousands of contacts with the wall while giving up energy to the cavity through the ($F = 1$, $m = 0$) $\rightarrow$ ($F = 0$, $m = 0$) transition. If this energy exceeds the cavity loss self-sustaining oscillations will result. The hydrogen maser has a short-term stability approaching 1 part in 10^{16} and about 50 hydrogen maser clocks now contribute to TAI.

See also **Clocks and Watches, Quartz; Global Positioning System (GPS)**

JAMES MCASLAN STEELE

Further Reading

Audoin, C. and Vanier, J. Atomic frequency standards and clocks. *J. Phys. E. Sci. Instrum.* 9, 697–720, 1976. Still an excellent review of the subject at some remove in time and the source of Figures 4, 5, and 6.

Audoin, C. and Guinot, B. *The Measurement of Time*, Cambridge University Press, Cambridge, 2001. Original French edition, *Les Fondements de la Mesure du Temps: Comment les Fréquences Reglent le Monde*, Masson, Paris, 1998. A physicist and an astronomer combine to provide a comprehensive survey of the field.

Forman, P. Atomichron®: the atomic clock from concept to commercial product. *Proc. Inst. Electr. Electr. Eng*, 73, 7, 1181–1204, 1985. A detailed and fascinating account of the genesis of the first commercial cesium clock.

Jones, T. *Splitting the Second: The Story of Atomic Time*. Institute of Physics Publishing, Bristol, 2000. Sponsored

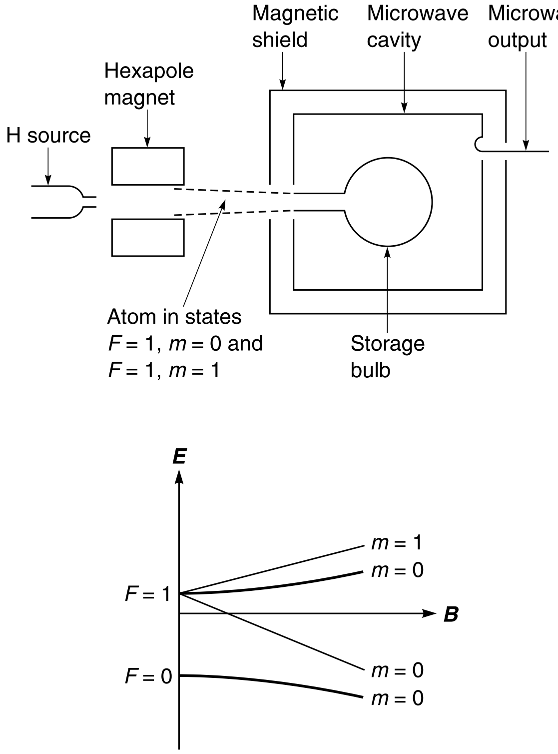


Figure 6. Atomic hydrogen maser.
[*Source: Audoin, C. and Vanier, J. Atomic frequency standards and clocks, J. Phy. E: Sci. Instrum., 9, 697–720, 1976.*]

by NPL; an excellent book for anyone interested in modern timekeeping.

Ramsey, N.F. *Molecular Beams*. Clarendon Press, Oxford, 1956. The classical reference work.

Ramsey, N.F. History of atomic clocks. *J. Res. Nat. Bureau of Stand.*, 88, 5, 301–320, 1983. An account of developments in some of which the author played a major role.

Vanier, J. and Audoin, C. *The Quantum Physics of Atomic Frequency Standards*, vols 1 and 11. Institute of Physics Publishing, Bristol, 1989. A magisterial contribution to the subject in over 2000 pages.

Cloning, Testing and Treatment Methods

Definitions

The use of the word "cloning" is fraught with confusion and inconsistency, and it is important at the outset of this discussion to offer definitional clarification. For instance, in the 1997 article by Ian Wilmut and colleagues announcing the birth of the first cloned adult vertebrate (a ewe, Dolly the sheep) from somatic cell nuclear transfer, the word clone or cloning was never used, and yet the announcement raised considerable disquiet about the prospect of cloned human beings. In a desire to avoid potentially negative forms of language, many prefer to substitute "cell expansion techniques" or "therapeutic cloning" for cloning. Cloning has been known for centuries as a horticultural propagation method: for example plants multiplied by grafting, budding, or cuttings do not differ genetically from the original plant. The term clone entered more common usage as a result of a speech in 1963 by J.B.S. Haldane based on his paper, "Biological possibilities for the human species of the next ten-thousand years." Notwithstanding these notes of caution, we can refer to a number of processes as cloning. At the close of the twentieth century, such techniques had not yet progressed to the ability to bring a cloned human to full development; however, the ability to clone cells from an adult human has potential to treat diseases.

DNA Cloning

Cloning may also refer to the process of creating identical segments of DNA. The development of DNA sequencing and recombinant DNA techniques in the 1970s meant a gene of interest could be located and extracted from an organism, its DNA strand cut and cloned, and spliced into another "transgenic" organism. To create many copies of the desired DNA segment for gene splicing or for DNA testing, the polymerase chain reaction (PCR), invented by Kary Mullis in 1985, is used. The PCR technique mimics the natural DNA replication that takes place during cell division.

Embryo Splitting (Blastomere Separation)

Embryo splitting is a method whereby the cells of a very early embryo (two-, four-, or eight-cell stage) are separated out to continue developing individually. Since at this stage each blastomere is "totipotent," the undifferentiated cell will continue to divide, then differentiate into functionally specific cell types, and eventually develop into a complete organism. Each of the resulting organisms will be genetically identical. This process of embryo splitting is, in many respects, a technically induced means identical to that by which identical (monozygotic) twins occur naturally. The technique was first demonstrated in 1902 by the German embryologist Hans Spemann who divided a two-celled salamander embryo into two. Each cell grew to adulthood.

In the context of human reproductive fertility, this form of cloning was first used experimentally in 1992 when researchers split the cells of 17 chromosomally abnormal human embryos to determine if development would continue. While

there was no intention to reimplant the embryos to develop *in vivo*, it nevertheless demonstrated the application of the technique in human reproduction.

There are, however, important differences between embryo splitting and the nuclear transfer method discussed below. The primary difference is that embryo splitting requires that an organism has already begun developing following fertilization in which the two separate gametes of parents are combined. The resulting split embryos are therefore "clones" of each other and not clones of either parent. Somatic nuclear transfer on the other hand can be used to produce a genetic replica of an already known adult being.

Nuclear Transfer

The concept of nuclear transfer was first proposed by Spemann in 1936 who suggested that it may be possible to take the nucleus of a cell, containing the full complement of genes necessary for development, and transfer it into an ovum from which the nucleus had been removed.

However, it was not until 1952 that the principle was demonstrated by Robert Briggs and Thomas J. King when they applied nuclear transfer to frogs. In their subsequent experiments throughout the 1950s and 1960s, they found that the procedure was only successful where the nuclei was taken from early embryonic undifferentiated stem cells. The use of adult and differentiated cells on the other hand resulted in very rare successes, where resulting tadpoles developed abnormally, if at all. This suggested that it was not possible to reverse the specialized differentiation of cell nuclei such that they could be induced to develop successfully into a complete adult organism.

However, the early 1960s experiments by John Gurdon seemed to suggest that adult differentiated cells taken from the intestines of South African frogs could be used to produce clones. The claim was subsequently contested on the basis that undifferentiated stem cells, which are naturally present in adult tissues, may have been mistakenly used instead.

Similar work in mammals has progressed far more slowly than in amphibians. In 1984, the Danish scientist Steen Willadsen successfully applied nuclear transfer to sheep for the first time by deriving nuclei from early embryo cells (stem cells). The next key development in the field occurred in 1996 when Ian Wilmut's team at the U.K.'s Roslin Institute announced the birth of Dolly, claiming that the sheep had been cloned using the nuclei of an adult somatic cell rather than that of an embryonic stem cell. The nuclei had been treated to reverse the process of cell specialization. The technical shift from embryonic to somatic nuclear transfer holds the possibility of much greater flexibility in the production of cloned and transgenic mammals. By using cultured (adult) somatic cells, researchers need not be overly restricted by having to rely on a limited number of available human embryos (from aborted fetuses or unused embryos from fertility treatment).

Cloning for Therapeutic Applications

International policymaking in the late 1990s sought to distinguish between the different end uses for somatic cell nuclear transfer resulting in the widespread adoption of the distinction between "reproductive" and "therapeutic" cloning. The function of the distinction has been to permit the use (in some countries) of the technique to generate potentially beneficial therapeutic applications from embryonic stem cell technology whilst prohibiting its use in human reproduction. In therapeutic applications, nuclear transfer from a patient's cells into an enucleated ovum is used to create genetically identical embryos that would be grown *in vitro* but not be allowed to continue developing to become a human being. The resulting cloned embryos could be used as a source from which to produce stem cells that can then be induced to specialize into the specific type of tissue required by the patient (such as skin for burns victims, brain neuron cells for Parkinson's disease sufferers, or pancreatic cells for diabetics). The rationale is that because the original nuclear material is derived from a patient's adult tissue, the risks of rejection of such cells by the immune system are reduced.

See also **Fertility, Human; Genetic Engineering, Methods**

NIK BROWN

Further Reading

Kolata, G. *Clone: The Road to Dolly and the Path Ahead.* Penguin, London, 1997.

Lauritzen, P., Ed. *Cloning and the Future of Human Embryo Research*. Oxford University Press, Oxford, 2001.

Wilmut, I., Schnieke, A.E., McWhir, J. *et al.* Viable offspring derived from foetal and adult mammalian cells. *Nature*, 385, 810–813, 1997.

Wilmut, I. Cloning for medicine. *Scientific American*, 279, 6, 58–63, 1998.

Coatings, Pigments and Paints

The development and application of pigments, paints, and coatings have been an integral part of human development from Paleolithic cave paintings, the art of early civilizations, and protection of buildings from rain. During the twentieth century, understanding of chemicals and the manufacturing need for high-quality decorative and protective coatings drove rapid progression in paint technology. All paints employ the same basic ingredients: pigment to provide color; a medium to bind or suspend the pigment, including emulsions such as resins or oils; and a solvent carrier, which acts to wet the surface to ensure adhesion and thins the resin to make it easy to apply.

Early pigments such as those used in Minoan frescoes and Anasazi rock art and the use of henna as body paint originated primarily from natural sources. Clays, mineral pigments such as iron or chromium oxide, vegetable dyes, animal sources such as shells or urine, as well as precious metals and gems gave rise to a broad selection of pigments that were often unique to one region (e.g., lapis lazuli) and thus highly prized as trade items. But value wasn't limited merely to pigments. Preservation of painted surfaces required protective resins. The most successful early coating was lacquer, used in China since at least the 1300 BC. Lacquer was processed—using a highly guarded secret formula—from resin from the lacquer tree (*Rhus vernicflua*). Shellac, produced from the gum secreted by an insect native to India and southern Asia, makes a varnish when mixed with acetone or alcohol and was used from the eighteenth century. Natural plant resins dissolved in oil or solvent such as turpentine were used in the nineteenth century—evaporation of the solvent leaves a lacquer coating.

Although mineral oxides were often naturally found in soft clays or chalks, binders are needed to enable the pigment to disperse over the surface evenly. Traditional binders such as gum arabic and other natural resins, beeswax, glycerin, egg, animal glue, and linseed oil have been used for centuries. Pigment–oil mixtures enhanced the pigment, increasing brilliance, translucence, and intensity of color. By the early twentieth century, chemists understood more about the physical properties of different pigments, providing them with more strategic approaches to improving binders and carriers.

By the mid-nineteenth century, use of synthetic pigments and dyes as alternatives to natural plant or mineral pigments had increased. A wide range of cadmium yellows and reds were soon available. By 1930 titanium dioxide, a bright white pigment with great covering power derived from titanium ore, was successfully applied to paint. Its fine particles were easily suspended in linseed oil.

Until the mid- to late 1800s paint was generally prepared onsite by painters themselves. Powders could be mixed with water (as in distemper or whitewash) or white lead paste was mixed with linseed oil. Hand-ground pigments were often coarse, and dispersal by hand resulted in unevenness of color. In 1867, D.R. Averill of Ohio patented the first ready-mixed paints in the U.S., but did not market the paints. Mass produced ready-mixed paints were developed by the Sherwin-Williams Company by 1880 and sold in tins in a wide range of colors.

By the early twentieth century, in the quest to make more durable paints and coatings for a growing manufacturing industry, cellulose became a critical ingredient in synthesized lacquer. Cellulose acetate, which is produced from a cellulose source such as paper or cotton, is mixed with acids and dispersed in solvents to form a low-viscosity lacquer. By 1928 the development of a nitrocellulose solution with a dense pigment base and low viscosity became the paint of choice for the mass production of planes and automobiles, coinciding with the start of spray application.

Research in synthetic resins and polymers led to the use of alkyd and amino resins in coatings and varnishes from the 1920s. Alkyd resins, derived from glycerol processed from animal and vegetable fats, were introduced in paints in 1927. Durable, flexible, inexpensive to produce, and with strong adhesion to most surfaces, alkyd resins were popular as industrial coatings, for example glossy finishes in the automotive industry. Urea and melamine formaldehydes were used from the 1930s in industrial and decorative laminates.

World War II stimulated discoveries in the paint industry. By the end of the war, linseed oil and the solvents that were used to cut it became scarce. Thus from the 1940s and 1950s alkyd resins began to replace linseed oil as a binder in paints. Wartime scarcity of rubber also stimulated research into synthetic latex, and a byproduct was the introduction of latex as a waterborne binder in paints. The first synthetic latex (styrene-butadiene) emulsion paints were introduced for architectural use in 1948, and were also quickly adopted, replacing oil and solvent-based paints. By the 1960s other polymers, vinyl acetate and methyl methacrylate (acrylics) gradually replaced styrene-butadiene latex as paint binders, becoming widely used in household paints. Acrylic had been used as a fast-

drying automobile lacquer from the 1950s, replacing nitrocellulose lacquers, and from the 1960s was also used in automobile enamel paints. The acrylic enamel paints had drying times competitive with lacquers, improved gloss and durability, and improved resistance to ultraviolet (UV) damage. Powder coatings, developed in the 1960s as an alternative to liquid paints, are applied dry as an acrylic, silicone, or polyester resin powder and form a coating on heating by cross-linking of polymers. From the 1990s teflon (polytetrafluoroethylene) was used as a painted sealant on aircraft and automobiles.

In the 1980s polyurethane paint binders and clear varnishes evolved, using a polymer formed by reacting an isocyanate group with another group, often a hydroxyl group. An alkyd resin reacted with a polyisocyanate forms the common polyurethane varnish used in residential applications. In industrial applications, the coating reacts and cross-links in one of several ways: on application and exposure to moisture from the air, by heat curing, by mixing with a catalyst, or by heat activation of a latent catalyst. Cross-linking improves the toughness of the resin.

By the 1960s, coatings and paints were driven by chemistry, with new materials created by research groups answering the need for better, stronger, and cheaper coatings for plastic, paper, and of course, wood and metal. Formaldehyde coatings are also widely used for permanent press fabrics. Paints continue to evolve, with non-glare, UV-resistant paint, paint that deadens sound, and acrylics that absorb stains.

Environmental Concerns

Lead paint is durable, cheap to produce, and widely used as interior and exterior paint, despite its known toxicity. Industry consensus standards limiting the use of lead pigments date back to the 1950s, when titanium dioxide started to replace lead as the white pigment of choice. Use of lead in household paints in the U.S. was banned by the Consumer Product Safety Commission in 1978.

In response to the toxic properties of solvents used in varnishes, lacquers, and urethane, paint manufacturers in the 1990s shifted toward producing low VOC (volatile organic compound) paints. Today, alkyds (which can be made soluble in water) have replaced most oil-based emulsifiers, and latex has replaced solvent-cut paints. The waterborne products have disadvantages such as lower gloss, and are also more expensive to produce. Some high-solid products were developed, where some of the VOC was replaced by lower molecular-weight resins, which did not increase the viscosity. Powder coatings have near-zero VOC.

See also **Dyes; Synthetic Resins**

LOLLY MERRELL

Further Reading

Ball, P. *Bright Earth: Art and the Invention of Color*. Farrar, Straus & Giroux, New York, 2002.

Gettens, R.J. and Stout, G.L. *Painting Materials, a Short Encyclopedia*. Dover Publications, New York.

Hammond, K. A brief history of paints. *Polym. Paint Col. J.*, 193, 24–25, 2003.

Marrion, A., Ed. *The Chemistry and Physics of Coatings*. Royal Society of Chemistry, Cambridge, 1994.

Coherers and Magnetic Detectors, *see* **Radio Receivers, Coherers and Magnetic Methods**

Color Photography

Most colors in nature can be represented as a mixture of the primary colors of red, green, and blue, and almost all photographic color processes begin by recording the proportions of these colors captured from the original scene. Two different paths may be followed. One involves adding together the three primary colors to build an image—an additive process; the other starts with pure white light containing all primary colors and removing those not wanted by superimposing positive images in complementary colors: cyan, magenta, and yellow—a subtractive process.

The possibility of producing photographic images in natural colors was a dream of the pioneers of photography, but for decades the only means available was to hand color with paint or powder. As late as 1882 the English photographer, John Werge, claimed that there was "little or no probability of polychromatic pictures being obtained in the camera direct." Nevertheless, it was a series of nineteenth century discoveries and innovations that were to make color photography possible. The single most influential work was that of the Scottish physicist, James Clerk Maxwell. In 1861 he photographed a tartan ribbon through separate blue, green, and red filters to make three positive lantern slides. Using blue, green, and red light sources in a lantern, the projected images were superimposed on the screen to produce a colored picture approximating to the color of the original tartan. By the early 1890s, the American, Frederick Ives, had devised a color photography system based on Maxwell's principles. The same

decade also saw the first screen plate color photographs, made by exposing a photographic emulsion through a mosaic of tiny primary color filters. These primitive processes were unsatisfactory in many respects however and color photography only became a reality in the twentieth century.

In 1904 the Lumiere brothers Auguste and Louis announced a new screen plate process prepared by a randomly scattered mixture of starch grains dyed in the three primary colors and coated with a standard panchromatic emulsion. Each grain acted as a tiny filter so that from a single exposure in a camera, three "separations" were made on the one plate. After processing, the original subject was reproduced as a glass transparency in natural colors. The process required long exposures and the final images were dense and needed bright illumination for viewing. However, it was the first fully practicable color process. When marketed in 1907 as the Lumiere autochrome process, it soon became popular with both professionals and amateurs. It received an early boost in America when Edward Steichen exhibited examples along with specimens by other notable photographers. Other screen plate processes soon followed. The Thames Plate of 1908 and the Paget process of 1914, both with regular patterned screens, required shorter exposures than the Autochrome process and enjoyed success in England. With the rising popularity of roll film cameras, the screen plate principle was transferred to celluloid film. Agfacolor and Dufaycolor roll film was widely sold throughout Europe in the 1930s.

The screen plate processes described were additive processes, but the other major means of producing color images during the first 40 years of the twentieth century produced color by subtraction. Ingeniously designed three-color cameras containing mirrors or beam splitting devices were used to produce the color records in the form of separation negatives. Cyan, magenta and yellow carbon prints were then made from those negatives. When superimposed in exact register a print in full natural color was produced. A carbon print consists of hardened gelatin containing colored pigment and is completely stable. Its great advantage at the time was that richly colored images could be produced on paper. Several color processes based on refinements of the carbon process were marketed in Europe and America. Perhaps the most successful was the Carbro (or Ozobrome) process, which derived from H. E. Farmer's 1889 discovery that bichromated gelatin in contact with a silver image becomes insoluble in water without the action of light. In the 1920s and 1930s the Autotype Company in England successfully marketed the process worldwide as Trichrome Carbro with the slogan "no daylight required." Unlike some carbon color processes, it was comparatively easy to work and became popular with amateur photographers.

Early in the twentieth century, it seemed that color photography might be achieved by coating three different color sensitive emulsion layers onto a single film base to form a tripack. In 1912 a German, Rudolph Fischer, patented a tripack containing color-forming materials within each emulsion layer but formidable technical problems prevented commercial exploitation. Over 20 years passed before two musicians working for Eastman Kodak, Leopold Godowski and Leopold Mannes, devised a slightly different tripack system involving three thin monochrome emulsion layers, each sensitized to a different color. Unlike Fischer's technique, color couplers were not introduced until processing, which was complex and included a three-stage development followed by bleaching to remove unwanted dyes. Kodak marketed the new process on roll film, for cine cameras in 1935, and for still cameras a year later. Called Kodachrome, it was received enthusiastically by the public and marks the beginning of modern subtractive color photography.

The German Company, Agfa, was also working on tripack film and in 1936 finally introduced a commercial development of Fischer's integral tripack. It was again constructed on the three-layer principle with the important difference that the color couplers and dyes were incorporated within the emulsion layers. The Agfa system was much simpler to process and Kodak soon developed its own version, a negative–positive process introduced in the early 1940s as Kodacolor, the precursor of almost all current color negative film. A variant, Ektachrome, for color transparencies followed, although an improved form of Kodak's original Kodachrome was still made available for high-quality color slides. Also developed during the 1940s were papers coated with a three-layer emulsion incorporating dye-releasing color couplers, which allowed the production of good quality color prints (sometimes termed C-type prints). Modern improved versions are used today to produce the millions of images processed in mini labs and shops throughout the world.

See also **Cameras, 35 mm; Cameras, Automatic**

John Ward

Further Reading

Coe, B. *Colour Photography*. Ash & Grant, London, 1978.
Friedman, J.S. *History of Color Photography*. American Photographic, Boston, 1944. Reprinted by Focal Press, London, 1968.
Hunt, R.W.G. *The Reproduction of Colour*. Fountain Press, Watford, 1975.
Mees, C.E.K. *From Dry Plate to Ektachrome Film*. Eastman Kodak, New York, 1961.
Sipley, L.W. *A Half Century of Color*. Macmillan, New York, 1951.
Spencer, D.A. *Colour Photography in Practice*. Pitman Greenwood, London, 1938.
Spencer, D.A., Ed. The Focal Dictionary of Photographic Technology. Focal Press, London, 1973.
Wall, E.J. *The History of Three-Color Photography*. American Photographic, Boston, 1925. Reprinted by Focal Press, London, 1970.

Combinatorial Chemistry

Combinatorial chemistry is a term created about 1990 to describe the rapid generation of multitudes of chemical structures with the main focus on discovering new drugs. In combinatorial chemistry the chemist should perform at least one step of the synthesis in combinatorial fashion. In the classical chemical synthesis, one synthetic vessel (flask, reactor) is used to perform chemical reaction designed to create one chemical entity. Combinatorial techniques utilize the fact that several operations of the synthesis can be performed simultaneously.

Historically, the first papers bringing the world's attention to combinatorial chemistry were published in 1991, but none of these papers used the term combinatorial chemistry. Interestingly, they were not the first papers describing the techniques for preparation of compound mixtures for biological evaluation. Previously, H. Mario Geysen's lab had prepared mixtures of peptides for identification of antibody ligands in 1986. Other laboratories heavily engaged in synthesizing multitudes or mixtures of peptides were Richard A. Houghten's laboratory in San Diego and Árpád Furka's laboratory in Budapest. The recollections of the authors of these historical papers were published in the journal dedicated to combinatorial chemistry, *Journal of Combinatorial Chemistry*.

Since the goal of combinatorial chemistry is the discovery of new compounds with interesting properties (biological as new pharmaceuticals or physical as new materials), the chemists want not only to make as many compounds as possible as quickly as possible, but to make these compounds as different from each other as possible in order to cover what is referred to as chemical space. This space stretches over all theoretically possible structures and conformations of all compounds within a given range of size. When the structures of a set of compounds made by combinatorial chemistry are evenly distributed over the respective chemical space, one of the compounds has a better statistical chance of being identical or at least similar to the "optimal" structure (conformation) for a desired property (e.g., biological activity), as compared to a set of compounds that cover only a fraction of the chemical space.

A large set of related synthetic compounds is typically called a library (or combinatorial library). Libraries range in complexity from a few dozen up to millions of compounds. The central feature of combinatorial libraries is that all compounds making up the library represent combinations of two or more "building blocks" which are connected by chemical reactions.

Combinatorial libraries can be classified based on their composition or synthetic history (see Figure 7). Peptide-like (oligomer) libraries are composed of repeated units of similar building blocks connected by repetition of the same (or similar) chemical reaction. Glucose-like (scaffolded) libraries are based on the multifunctional scaffold, the functional groups of which are selectively employed in attachment of various building blocks. Benzodiazepine-like (condensed) libraries are created by connecting building blocks capable of forming unique structures depending on the order of performed reactions, where original building blocks may not be readily identifiable within the resulting library structure (various strategies and building block types can be used for forming the same resulting structures). Libraries can be structurally homogeneous or heterogeneous; that is, the compounds can have identical or variable "scaffolds" or "backbone." All combinatorial libraries can also be complete (containing all theoretically possible combinations of used building blocks), or incomplete (containing only a fraction of all possible compounds). A complete combinatorial library is composed of all possible permutations of the building blocks at their respective positions. If the scaffold of a library has three attachment points (prospective diversity positions), and ten different building blocks are used for each diversity positions, then the complete combinatorial library is composed of $10^3 = 1000$ compounds.

Synthesizing a combinatorial library can be rather straightforward, as the same protocol is typically used for all compounds, so that the synthesis method has to be worked out only

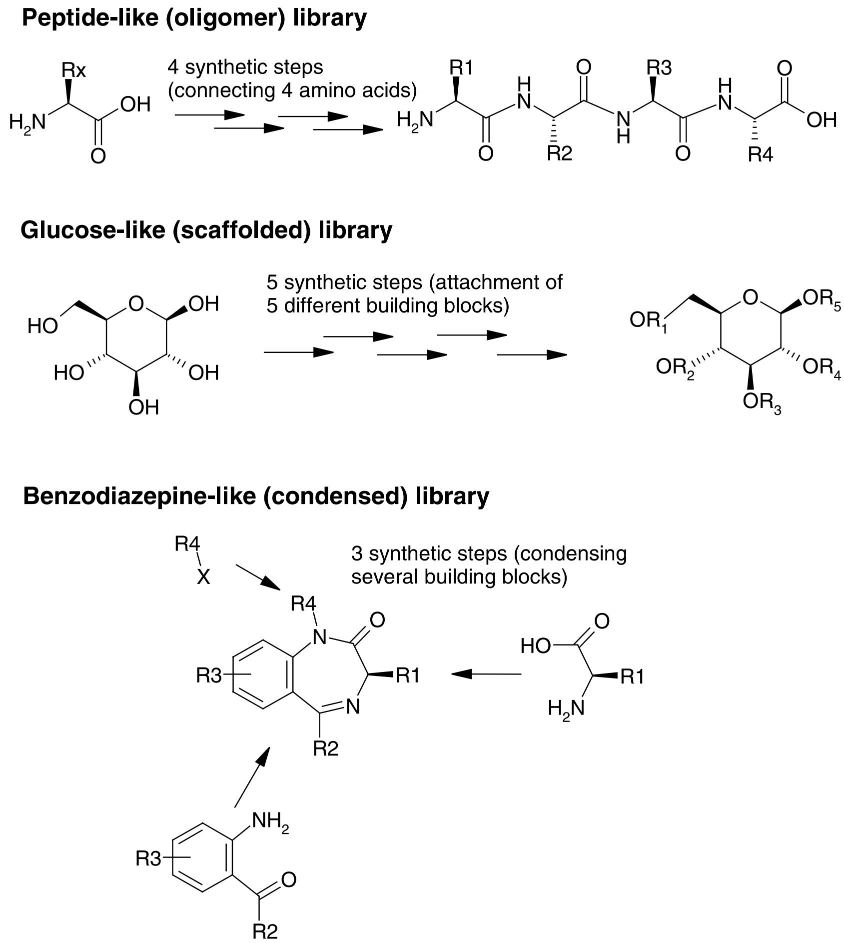


Figure 7. Different library types—peptide-like (oligomer) libraries, glucose-like (scaffolded) libraries, and benzodiazepine-like (condensed) libraries. Oligomeric libraries are built by connecting similar building blocks by repetition of one (or several) reactions—peptides and oligonucleotides are typical examples. Scaffolded libraries are constructed by modification of individual functional groups on the template (scaffold) molecule. In condensed libraries it may be difficult to trace the character of building blocks used for their construction.

once. This, however, is not always as easy as it may sound (with issues covering the choice of synthetic strategy, in solution or on solid support, chemistry of attachment of the first building block, protection and deprotection strategies, release from solid support, etc.), as the optimal reaction conditions can vary greatly among the different building blocks used for a particular step.

An important prerequisite for combinatorial chemistry is the availability of methods for parallel synthesis. Prototypical of combinatorial chemistry techniques is Houghten's "tea bag" method. In this technique the solid support (functionalized polystyrene resin) is sealed in packets made of polypropylene mesh (Figure 8), which is permeable for solvents and reagent solutions. Up to several hundreds of such resin packets can be processed simultaneously in common reaction vessels. After each step the packets are resorted for the next synthesis step. Resorting is either based on readable alphanumeric labels, or it can be simplified by enclosing a radio-frequency tag in the tea bag. In this way up to a thousand compounds can be synthesized using a reasonable number of reaction vessels. For example in the peptide synthesis, only 20 reactors with individual amino acids are required for the synthesis of basically unlimited number of (natural) peptides of any length in parallel.

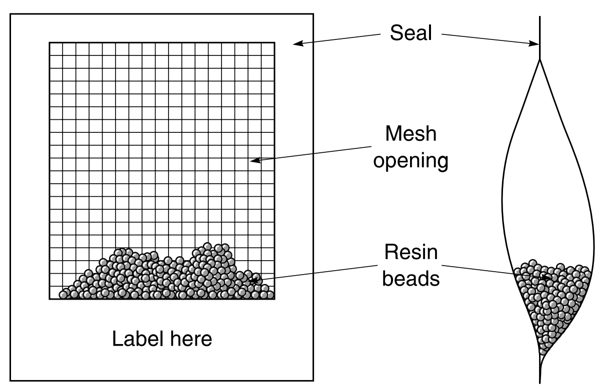


Figure 8. Schematic drawing of the "tea-bag."

A powerful, yet simple method for manual or semiautomated solid-phase synthesis of mixtures of up to millions of compounds is the ''one-bead–one-compound'' approach (Figure 9). It has also been referred to as ''split-and-mix'' or ''divide–couple–recombine'' approach, and is based on coupling each building block to separate portions of the solid-phase resin, followed by combining and mixing all resin portions, before dividing the resin again for the next synthesis step. By repeating this procedure three more times, and using 20 different building blocks for each synthesis step, a library of 160,000 (20^4) compounds can be readily prepared. This process yields libraries containing an individual, unique compound on each resin bead. After assembling the library on the resin, it can be either cleaved for bioassays in solution, or left on the resin for solid-phase assays. The bio-assays are typically performed on single beads, so that the screening format of one-bead–one-compound libraries is that of single compounds, rather than compound mixtures. The one-bead–one-compound library principle is based on the statistical distribution of the particles in the process. However, this statistical nature of the process can be eliminated by the use of continuously divideable solid supports, such as membranes or threads. In this case, all members of the library are guaranteed to be prepared, and none of them is prepared in more than one copy.

MICHAL LEBL

Further Reading

Frank, R. and Doring, R. Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper discs as segmental solid supports. *Tetrahedron*, 44, 19, 6031–6040, 1988.

Furka, A., Sebestyen, F., Asgedom, M. and Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Peptide Prot. Res.* 37, 487–493, 1991.

Geysen, H.M., Meloen, R.H. and Barteling, S.J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci. USA*, 81, 3998–4002, 1984.

Houghten, R.A. General method for the rapid solid-phase synthesis of large numbers of peptides: Specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. USA*, 82, 5131–1315, 1985.

Houghten, R.A., Pinilla, C., Appel, J.R., Blondelle, S.E., Dooley, C.T., Eichler, J., Nefzi, A. and Ostresh, J.M. Mixture-based synthetic combinatorial libraries. *J. Med. Chem.*, 42, 19, 3743–3778, 1999.

Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. and Knapp, R.J. A new type of synthetic peptide library for identifying ligand- binding activity. *Nature*, 354, 82–84, 1991.

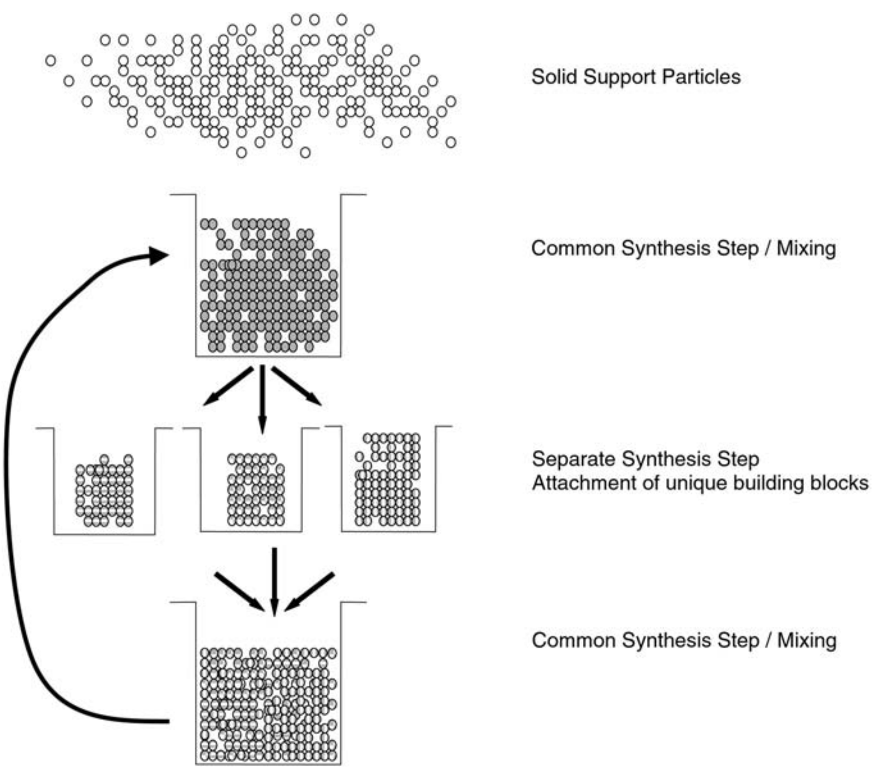


Figure 9. Principle of one-bead–one-compound library synthesis. Resin particles are exposed to only one reagent at a time and therefore each particle can contain only one structure. The process of separating the solid support into aliquots and mixing them is repeated as many times as there are steps of the library building using the unique building blocks. A multistep synthesis of a library, in which three of the steps use the various building blocks (10 different building blocks are used in each step), would generate $10 \times 10 \times 10 = 1000$ different bead populations. If 1 gram of 130-micrometer polystyrene beads were used for the synthesis (1,000,000 beads), there will be in average 1,000 beads carrying the same compound. If only 1,000 beads are used for the synthesis (very unlikely), the chance of having any particular structure represented in the library would be only about 70 percent.

Lebl, M. Parallel personal comments on "classical" papers in combinatorial chemistry. *J. Combinatorial. Chem.*, 1, 1, 3–24, 1999.
Nicolaou, K.C., Xiao, X.Y., Parandoosh, Z., Senyei,A. and Nova, M.P. Radiofrequency encoded combinatorial chemistry. *Angew. Chem. Int. Ed.* 34, 2289–2291, 1995.
Stankova, M., Wade, S., Lam, K.S. and Lebl, M. Synthesis of combinatorial libraries with only one representation of each structure. *Peptide Res.* 7, 6, 292–298, 1994.

Useful Websites

Compilation of papers in molecular diversity field, 1996: http://www.5z.com.

Communications

To the many technological inventions of the nineteenth century that improved communications, such as the telegraph and the telephone, the twentieth century added motion pictures, radio, television, and the Internet. Most of the products created for improving communication in the nineteenth century improved communication between individuals. During the twentieth century, new technologies added ways for groups or organizations to communicate to other groups, marking the birth of mass communications. These new media would have profound effects on entertainment, how people received news, and politics.

Motion Pictures

The first major communication improvement to be commercialized in the twentieth century was the motion picture. Thomas Edison invented the first practical motion picture camera in the U.S., and in 1896 he showed a motion picture to the public in the New York City Music Hall. The U.S. was the pacesetter for the film industry and New York City was the early center of the motion picture business. The first narrative film was *The Great Train Robbery* (1903). During World War I, Hollywood began to replace New York as the home of the movie industry. By the 1920s, Hollywood had clearly become the movie-making capital of the world and silent-film stars such as Charlie Chaplin, Buster Keaton, and Mary Pickford established themselves there.

Technological developments in the early twentieth century made sound motion pictures possible and greatly changed the film industry. In 1926, Warner Brothers, then a relatively minor studio, released the film *Don Juan* with a synchronized orchestral accompaniment. The studio had purchased the sound-on-disk Vitaphone system from American Telephone and Telegraph (AT&T). Warner Brothers sought to make short-term profits by supplying the technology to theaters that could not afford to hire live orchestras. This first attempt was successful enough that the first talking movie—*The Jazz Singer* staring Al Jolson—was released in 1927. In addition to the orchestral accompaniment, this film also featured popular songs and dialogue. The new sound films enjoyed great success and almost all Hollywood films included sound by the late 1920s, leading to greatly increased profits for the studios.

The introduction of sound did lead to some problems and changes in the film industry. Problems included the tremendous expense now involved in the production of motion pictures and the primitive nature of microphones, which forced actors to remain almost stationary and would pick up the sound of cameras and other set noise. Changes included the replacement of many silent-era actors with new actors with stage experience due to the fact that with sound, film actors had to have pleasant sounding voices without strong foreign accents. Sound led to the production of more realistic films, including crime epics and historical biographies. Musicals also became an important film genre, including the animated musicals of Walt Disney.

Movies would become the world's leading form of entertainment until the advent of television after World War II. For many, films represented the decadence of twentieth century society, displaying modern sexual mores on the screen. Along with radio, the film industry also became big business. Furthermore, the motion picture industry centered in Hollywood served to export the culture of the U.S., as moviegoers around the world viewed U.S.-made films.

Although motion pictures were mostly used for entertainment, showings often included such current news events as wars, parades, and speeches. In the 1930s, after sound had been added, newsreels covering the week's major events were shown in most theaters along with the movies. Authoritarian governments that emerged between the two world wars became particularly adept at utilizing film as a propaganda tool. The Bolsheviks in the Soviet Union and the Nazis in Germany successfully used motion pictures for political ends. The most famous example of this political use is *Triumph of the Will*, a 1934 film by the German filmmaker Leni Riefenstahl that depicted a Nazi rally at Nuremberg.

During the war, film was used to support the war aims of governments. In the U.S., for example,

Hollywood backed the government's information campaign through the Bureau of Motion Picture Affairs, which produced commercial features with patriotic themes. Hollywood also produced documentaries such as Frank Capra's *Why We Fight*, which sought to explain the war to both soldiers and civilians.

Immediately after the war, Hollywood enjoyed a brief boom, as two-thirds of Americans went to the movies at least once a week. Soon, however, antitrust legislation, protectionist quotas abroad, and the rise of television cut into Hollywood's profits. The Cold War also greatly affected the film industry, as many suspected communists were blacklisted by the studios and film-making became more conservative. Traditional genres such as musicals and westerns continued after the war, while others grew in importance, including many lower-budget films that dealt with social problems such as racism or alcoholism. Also popular was *film noir*, which offered a dark interpretation of American society.

World War II devastated the film industries in much of Europe, the Soviet Union, and Japan. A postwar renaissance was led by Italy and its neorealist movement that attempted to show the reality of a country afflicted by warfare. Great Britain and France soon followed in reviving their film industries. Japan also was able to restore its motion picture industry after the war, as many studios were left intact. Akira Kurosawa led the Japanese revival with numerous films, including *Rashoman*.

A film industry developed in many Third World countries. India had a vibrant film industry led by director Satyajit Ray. Many of India's films provided an alternative cinema with artistic merit. At the same time, India also became the world's largest producer of low-quality films for domestic consumption, making more than 700 motion pictures in sixteen languages each year. Film was often the only access to audiovisual entertainment for the many poor and illiterate Indians.

Latin America and Africa also developed sometimes militant, alternative forms of film; for example in Cuba during the 1960s, when the country's revolution influenced world-renowned directors such as Tomás Gutíerrez Alea and Humberto Solás. The so-called *Cinema Nôvo* (New Cinema) developed in Brazil during the 1960s and spread to other Third World countries. While many Third World nations created sometimes revolutionary film genres, military dictatorships also repressed motion pictures in numerous countries such as Argentina.

Radio

By the early twentieth century, wireless communication began to appear. The first example of wireless communication was the radio. In 1895, Italian inventor Guglielmo Marconi transmitted the first wireless telegraph message. Starting in 1901, Marconi used radio telegrams to communicate with ships on the Atlantic Ocean. The usefulness of radio was seen in its use during the Russo–Japanese War in 1905. In the U.S., experimental broadcasting to a mass audience started in 1910 with a program by the famous singer Enrico Caruso at the Metropolitan Opera House in New York City. Perhaps the most dramatic example of radio's value in spreading information was its use in reporting on the sinking of the *Titanic* in 1912, which demonstrated radio's ability to allow people to experience distant events as they occurred. World War I interrupted some radio research, however the demands of military communications sped up the development of radio technology.

During the 1920s, what had been more of a hobby became a mass medium that played a central role in news reporting and entertainment. A number of experimental broadcasting stations had converted to commercial stations by broadcasting programs on a regular basis, including news such as the results of the 1920 presidential election in the U.S. In the U.S., because radio was a good way to communicate with large groups of people, broadcasting rapidly consolidated into national networks in order to attract advertising revenue to support news and entertainment programming. The Radio Corporation of America (RCA) created the first nationwide broadcast network, the National Broadcasting Company (NBC), in 1926. In Europe and some other parts of the world, governments generally controlled radio broadcasting.

Radio played an important role in twentieth century communications, as it allowed people much easier access to entertainment since many families owned radios. By the end of the 1920s, two-thirds of homes in the U.S. owned radio receivers. People no longer had to go to a concert, play, or sporting event to be entertained. Instead, they could now enjoy many forms of entertainment from the comfort of their own homes. Despite the fact that radio broadcasts could reach millions of people, the medium gave those in their homes a sense of immediacy and intimacy. Furthermore, unlike written forms of communication, no formal education was needed to enjoy radio programs. Many forms of popular entertainment shifted to the radio, allowing them to maintain and even

expand their audiences. Radio offered a wide variety of entertainment genres, including dramas, comedies, sports, and music.

Besides providing entertainment, supplying news, and making money for entrepreneurs, radio also proved to be an important tool for politicians, better enabling them to mobilize the masses. Perhaps best known are Franklin Roosevelt's "Fireside Chats," which allowed the president of the U.S. to reach the public directly during the Great Depression and World War II. As was the case with film, authoritarian regimes in particular made use of radio technology. Italy's Benito Mussolini pioneered the use of radio to address the nation. In the Soviet Union, the first experimental radio broadcasts began in 1919. In 1922, a central radio station in Moscow began broadcasting. By 1924, regular broadcasts could be heard throughout most of the USSR and by 1937, there were some 90 radio transmitters in operation in Stalin's Soviet Union. Leaders in Nazi Germany also made effective use of the radio during the 1930s and 1940s. In Japan, the right-wing government utilized radio to promote its goals leading up to World War II.

Radio has also become an important means of communications in other parts of the world. In Latin America, for example, radio, along with television, is the main medium for transmitting information. In most Latin American countries, radio reaches far more people than print media, due to lower rates of literacy and lack of purchasing power. From the 1930s to the 1960s, many radio stations broadcast *radionovelas*, serial radio programs similar to soap operas. Since the 1960s, such programming has largely moved to television. As was the case elsewhere, early Latin American radio also featured variety shows, dramas, sports, talk shows, and news.

Radio also contributed to the spread of Latin American culture to other parts of the world, especially in the realm of dance and music. Argentine tango, Mexican boleros, salsa from New York's Latin community, and Brazilian samba all became popular beyond the borders of Latin America in large part because of radio airtime. Samba, for example, emerged as a musical and dance form from the poor sections of Rio de Janeiro, the capital of Brazil at the time. From its Afro-Brazilian roots, samba emerged from a locally popular form to one that had a national importance in Brazil. As samba received increased radio airplay, it seemed to unite the country and came to represent Brazilian nationalism. Performers such as Carmen Miranda, who later also became a Hollywood film star, popularized the music on the Brazilian airwaves. Soon, listeners heard samba on their radios throughout the world, demonstrating that mass culture could spread from poorer countries to elite consumers around the globe due to communications technology such as radio.

From the 1990s, radio stations in Latin America have often become more specialized as they seek audiences. Amplitude modulation (AM) stations tend to carry news, talk, and local popular music. Also, they often cater to the interests of groups outside of the cultural and linguistic mainstream. For example, radio stations in Lima, Peru feature ethnic music and news in Quechua or Aymara languages for recent migrants from the highlands. Frequency modulation (FM) stations emphasize music, particularly national popular music or international music. International music tends to be popular among the young and affluent, while national music appeals to an older, more working-class audience.

Television

Another important twentieth century development was television, which would soon overshadow radio and motion pictures. In the U.S. during the late 1920s, many attempts were made to create an experimental telecast, and a few met with success, particularly RCA's efforts. In 1936, NBC provided 150 experimental television sets to homes in New York City and sent telecasts to them, the first show being the cartoon "Felix the Cat." By 1939, NBC was providing regular telecasts but to a limited market. When the U.S. entered World War II in 1941, however, all television projects were suspended until the war ended in 1945.

After the war, television development continued where it left off, with the invention of better television sets, creative programming, and larger markets. The first coast-to-coast program was President Harry Truman's opening speech at the Japanese Peace Treaty Conference in 1951. By the 1950s, television had become a profitable industry. Television enjoyed a "golden age" and increasingly replaced radio as the principal mass medium. Indeed, television became a key part of social life in the U.S. and other parts of the world. Following World War II, a growing number of people had more money and more leisure time, both of which were often spent on television.

While early televisions in the U.S. were largely affordable, they were often unreliable. Technological improvements soon made television much more reliable and appealing. These improve-

ments included the replacement of vacuum tubes with the transistor and the development of color sets. In 1953, the first color telecast was made, which spread so fast that by the 1960s, most telecasts were in color. Later advancements include the spread of cable television in the 1980s, which gave viewers access to dozens of specialized channels and challenged the power of the traditional television networks. Many of the newly available cable channels, such as MTV and CNN, would have important effects on society and culture. The end of the twentieth century witnessed the rise of satellite and high-definition television, which offered viewers even more choices and improved the technical quality of television.

Television continued the process of the globalization of U.S. culture, as viewers around the world watched comedies and dramas produced in the U.S. Sporting events also helped to spread the U.S.'s cultural values. The National Basketball Association (NBA) was particularly successful in its international marketing efforts, popularizing its sport around the globe and creating stars such as Michael Jordan, who arguably became the most recognized athlete in the world. In addition, U.S.-based businesses, such as Nike, benefited from the globalization of basketball through television, as the sport helped to sell more of its athletic shoes. Yet it was not only basketball and the U.S. that dominated the use of television. During the 1986 soccer World Cup in Mexico, games were played under the midday sun in order to be broadcast during primetime in European countries.

Television grew more slowly in the Soviet Union than in the U.S. and Western Europe. As late as 1960, only five percent of the Soviet population could watch television. Television audiences grew during the 1970s and 1980s, often at the expense of film and theater audiences. By 1991, 97 percent of the population could view television, and a typical audience for the nightly news from Moscow numbered 150 million.

Television also became available in Latin American countries during the 1950s, when it was largely restricted to an upper- and middle-class urban audience. In this early phase, programming was limited to live, local productions. From the 1960s, television became much more of a mass medium. In this period, much of the programming was imported from abroad, especially the U.S. By the 1970s and 1980s, high-quality national production appeared, especially in Brazil, Colombia, Mexico, and Venezuela. The most important and successful productions were *telenovelas*, a form of the soap opera. By 2000, in some countries, such as Brazil and Mexico, perhaps 90 percent of the population had regular access to television. While the figure is lower in rural areas, even this began to change in the 1980s when satellite dishes linked to repeater transmitters allowed for increased access in remote areas.

Computers and the Internet

While film, radio, and television all had dramatic effects on communications in the twentieth century, all three were still "one-way" media that lacked any sort of interactive capabilities. The advent of the computer revolution and in particular the Internet changed this situation in the late twentieth century. While early computers had been large and slow, by the 1970s and 1980s, engineers centered in California's so-called Silicon Valley created increasingly smaller computers with greater memory capacity. After these hardware developments, improvements in software followed that allowed computer users to word process, play games, and run businesses. These technological improvements in computer hardware and software would soon have a profound effect on communications and commerce with the development of the Internet.

The creation of the Internet was the result of attempts to connect research networks in the U.S. and Europe. In the 1960s, the U.S. Department of Defense created an open network to help academic, contract, and government employees communicate unclassified information related to defense work. After crucial technological advances in the 1970s, in 1980 the Department of Defense adopted the transmission control protocol/Internet protocol (TCP/IP) standard, which allowed networks to route and assemble data packets and also send data to its ultimate destination through a global addressing mechanism.

During the 1980s, the defense functions were removed from the network, and the National Science Foundation operated the remainder, adding many new features to the network and expanding its use around the world. While government agencies were the principal early users of the Internet, by the 1980s its use had spread to the scientific and academic community. By the 1990s, the Internet had become increasingly commercialized and privatized. The rise in the use of personal computers and the development of local area networks to connect these computers contributed to the expansion of the Internet. Starting in 1988, commercial electronic mail (e-mail) services were connected to the Internet, leading to a boom in

traffic. The creation of the World Wide Web and easy to use Web browsers made the Internet more accessible so that by the late 1990s, there were more than 10,000 Internet providers around the world with more than 350 million users.

In the early twenty-first century, the Internet is a critical component of the computer revolution, offering e-mail, chat rooms, access to the wealth of information on the Web, and many Internet-supported applications. The Internet has had a dramatic impact on global society. E-mail is rapidly replacing long-distance telephone calls, and chat rooms have created social groups dedicated to specific subjects, but with members living around the world. The Internet has not only changed how people communicate but also how they work, purchase, and play. Many people now work at home, using the Internet to stay in touch with the office. People have also begun to use the Internet for banking and shopping services rather than so-called "brick and mortar" locations.

The communications revolution of the twentieth century created many new social problems that will have to be addressed in the twenty-first century. While people have access to more information than ever before, that information, often unfiltered and invalidated, has created several generations of children who are seemingly immune to extreme violence. Health concerns are also an issue, as people spend less time in outdoor activities and more time sitting in front of the television or computer. The online nature of the Internet will also make privacy one of the major issues of the near future.

See also **Entertainment in the Home; Film and Cinema: Early Sound Films; Internet; Radio, Early Transmissions; Television, Cable and Satellite**

RONALD YOUNG

Further Reading

Abbate, J. *Inventing the Internet*. MIT Press, Cambridge, MA, 1999.

Barnouw, E. *A History of Broadcasting in the United States*, 3 vols. Oxford University Press, New York, 1966–1970.

Beniger, J. *The Control Revolution*. Harvard University Press, Cambridge, MA, 1986.

Baughman, J.L. *The Republic of Mass Culture: Journalism, Filmmaking, and Broadcasting in America since 1941*. Johns Hopkins University Press, Baltimore, 1992.

Burns, R.W. *Communications: An International History of the Formative Years*. Institute of Electrical Engineers, London, 2003.

Carey, J.W. *Communication as Culture: Essays on Media and Society*. Unwin Hyman, Boston, 1988.

Cook, D. *A History of Narrative Film*. W.W. Norton, New York, 1990.

Crisell, A. *Understanding Radio*. Methuen, New York, 1986.

Crowley, D. and Heyer, P. *Communication in History: Technology, Culture, Society*. Longman, New York, 1991.

Knight, A. *The Liveliest Art: A Panoramic History of the Movies*, revised edn. Macmillan, New York, 1978.

Leiner, B.M., *et al.* The past and future history of the Internet. *ACM Comm.*, 40, 2, 102–108, 1997.

McLuhan, M. *Understanding Media*. Signet, New York, 1964.

Williams, R. *Television and Society*. Fontana, London, 1979.

Composite Materials

Composite materials are defined as those that contain two or more materials that have been bonded together. Wood is an example of a natural composite material. It is made from lignin—a natural resin—which is reinforced with cellulose fibers. For thousands of years, straw has been used to reinforce mud bricks, forming a two-phase composite, and more recently, reinforced concrete has been developed. Concrete contains a cement binder with a gravel reinforcement. By adding another reinforcing material (steel rebar), concrete becomes a three-phase composite. Metal matrix and metal or ceramic composites have now also been developed.

Perhaps the most familiar class of composite materials are polymer composites. These are a class of reinforced plastics in which fibers are used to reinforce a particular polymer matrix. Phenolic laminates made from phenol formaldehyde resin and paper were developed around 1912, finding a use as an electrical insulating material. Glass is the most common fiber used to reinforce a polymer matrix, forming fiber-glass or glass-reinforced plastic. Other more expensive and higher strength materials such as carbon and, more recently, aramid fibers, are used in advanced applications such as aircraft components. Polyester, vinyl ester and epoxy resins are the most commonly used thermoset resins for the formation of the polymer matrix. PEEK (polyether ether ketone) may also be selected as the resin for applications where cost is less of a problem, such as in aerospace applications. Other resins, including phenolic, silicone, and polyurethane may be used for particular functions. The designer can alter the chemical resistance properties, service temperature capabilities, weather resistance, electrical properties, resistance to fire and adhesive properties by choosing the right type of resin.

Fibers for reinforcing can be obtained in a variety of different formats. They can be woven or

multiaxial, continuous or chopped (to give a random form). Alternatively they can be variations or combinations of all these types. By selecting the right orientation of fiber lay-up, it is possible to "design-in" a variety of properties connected with the physical strength of the ultimate composite.

Composite materials are often what we call layered composites—they are made up of layers of fibers, supplied as plies or lamina. A single ply comprises fibers oriented in a single direction (unidirectional) or in two directions (bidirectional); for example, in a woven fabric. Random fiber layers are also used. These are supplied as "prepregs," meaning that they are prepared before being molded by using a thermoplastic binder that is applied to the reinforcement and then heating the reinforcement. This softens the binder, which can then be formed into the desired shape in a separate operation. A specialist form of composite production involves filament winding on a mandrel prior to molding, producing high-strength rods.

Polymers fall into two classes—thermoplastics and thermosetting plastics, or thermosets.

Thermoplastics include familiar plastics such as acrylic (polymethyl methacrylate), polyethylene, acrylic, polypropylene, and polystyrene. They can be heated and formed, then remelted and reformed into a new shape.

Composite materials are normally made with thermoset resins. These start as liquid polymers. Following the cross-linking process, which they undergo on heating, the liquid polymers are changed during the molding process into an irreversible solid form. Composite materials therefore have superior properties to thermoplastics as they have better heat and chemical resistance, enhanced physical properties, and more structural resilience.

The development of composite materials that possess a range of advantageous properties have inspired a range of new uses in areas ranging from global uses, such as in aerospace, transport (both by road and sea), and building applications, as well as domestic products—high-tech sporting equipment, electrical goods and office equipment.

Composites have various advantages including high strength, light weight, and some temperature resistance. Their high strength has enabled the design of composite materials that meet the taxing needs of a particular function, for example an aircraft nose cone, which needs high impact strength. A variety of resins and reinforcements (from random to woven) can be used to meet the exact physical and mechanical properties needed for a particular structure.

The light weight of composite together with their high strength properties can be designed to meet very demanding specifications, for example in Formula One racing cars or in aerospace uses. Composites can be used to produce the highest strength-to-weight ratio structures known, as for example when aramid fibers are embedded in an epoxy matrix. Due to these high strength-to-weight ratios, these composites contribute enormous weight savings to aerospace structures—a vital consideration when more weight means more fuel use. Other advantages of these composite materials are their high resistance to corrosion and fatigue.

More complicated composite materials known as composite hybrids have been developed. These are formed by adding another material such as glass or aramid fiber to the original carbon fiber and epoxy matrix. These extra materials improve mechanical properties such as impact resistance and fracture toughness. Glass-reinforced plastic can gain improved stiffness by adding carbon and epoxy to the matrix.

While most modern engineering composites are made from a thermosetting resin matrix reinforced with fibers, some advanced thermoplastic resins may also be used. Other composites, in particular phenolic composites, have filler reinforcements, which may be mineral or fibrous, or a mixture of both. Foams and honeycombs can also be utilized as cellular reinforcements to bestow stiffness together with a very light weight. An example of such a polymer composite was used in the rotor blades of the EH101 Westland helicopter. The only disadvantage of this type of reinforcement, apart from its high cost, is that if the surface is damaged and the honeycomb becomes wet, it may lose its mechanical properties.

Composite plastic materials are made using a range of processes. The original process was a hand lay-up process, which can be slow, time-consuming and expensive. New manufacturing methods include pultrusion, vacuum infusion, resin transfer molding (RTM), sheet molding compound (SMC), low-temperature curing prepregs and low-pressure molding compounds. These methods are being used in high tech areas such as aerospace, and RTM in particular looks as though it is developing into a very good value for money process.

Choosing composites for certain applications is straightforward in some cases, but in others, their selection will rely on parameters such as service life, production run, complexity of mold, cost savings in assemblage, and on the experience and skills of the designer in tapping into the ultimate potential of composites. Sometimes it is best to

use composites together with more traditional materials.

The development of polymer composites has been pushed forward by the needs of the aerospace industry but has revolutionized the design of furniture, allowing the design of organic forms, as well as the design of boats, and sporting equipment where, as for example in the case of tennis rackets and racing cars, their performance has radically exceeded expectations.

See also **Ceramic Materials; Synthetic Resins; Thermoplastics**

SUSAN MOSSMAN

Further Reading

Kelly, A., Ed. *Concise Encylopaedia of Composite Materials*. Pergamon Press, Oxford, 1989.

Parkyn, B. Fibre-reinforced composites. in *The Development of Plastics*, Mossman, S.T.I. and Morris, P.J.T., Eds. Royal Society of Chemistry, Cambridge, 1994, pp. 105–114.

Computer and Video Games

Interactive computer and video games were first developed in laboratories as the late-night amusements of computer programmers or independent projects of television engineers. Their formats include computer software; networked, multiplayer games on time-shared systems or servers; arcade consoles; home consoles connected to television sets; and handheld game machines.

The first experimental projects grew out of early work in computer graphics, artificial intelligence, television technology, hardware and software interface development, computer-aided education, and microelectronics. Important examples were Willy Higinbotham's oscilloscope-based "Tennis for Two" at the Brookhaven National Laboratory (1958); "Spacewar!," by Steve Russell, Alan Kotok, J. Martin Graetz and others at the Massachusetts Institute of Technology (1962); Ralph Baer's television-based tennis game for Sanders Associates (1966); several networked games from the PLATO (Programmed Logic for Automatic Teaching Operations) Project at the University of Illinois during the early 1970s; and "Adventure," by Will Crowther of Bolt, Beranek & Newman (1972), extended by Don Woods at Stanford University's Artificial Intelligence Laboratory (1976). The main lines of development during the 1970s and early 1980s were home video consoles, coin-operated arcade games, and computer software.

Spacewar! grew out of the new "hacker" culture of the Tech Model Railroad Club (TMRC) at MIT. Intended as a demonstration program for a new PDP-1 computer donated by Digital Equipment Corporation, Spacewar! allowed players to control spaceships depicted on accurate star maps via the equally new precision cathode-ray tube (CRT) display Type-30. They maneuvered their spaceships via novel control boxes to avoid obstacles and fire torpedoes at their opponents. The result was a popular game available on PDP computers distributed to U.S. computer science laboratories in the 1960s and 1970s, such as the University of Utah's strong program in computer graphics. Nolan Bushnell, a former Utah graduate student and amusement park employee, recognized Spacewar!'s potential as a commercial product. With Ted Dabney, his co-worker at Ampex Corporation in California, he created "Computer Space" (1971) for Nutting Associates; this was a coin-operated version of Spacewar! set in attractive arcade cabinets.

Ralph Baer independently pursued the idea of creating video game consoles attached to home television sets. In 1971, he received a U.S. patent for a "television gaming apparatus," soon followed by acquisition of rights to it by Magnavox and, in 1972, by production of the first home video console, the Magnavox Odyssey. Bushnell and Dabney had by then created a new company, Atari Corporation; joined by Al Alcorn, another Ampex alumnus, Atari shipped Alcorn and Bushnell's electronic ping-pong game, "Pong," as an arcade game in November of 1972. Joining forces with the Sears department store chain, Atari released a home version of Pong in 1975. The phenomenal success of Pong stimulated competition leading to improved home and arcade consoles. The equally successful Atari 2600 VCS (video computer system), released in 1977, provided more flexibility and encouraged the separate development of game software distributed on cartridges and the hardware platforms accepting these games, at least in the home market. Activision, founded in 1979 by four former Atari game designers, was the first company exclusively focused on game software.

By the late 1970s, "home computers," single-user general-purpose computers with microprocessors, provided a new platform for electronic entertainment. Apple Computer's Apple II (1977) and the IBM Personal Computer (1981) featured color graphics, flexible storage capacity, and a variety of input devices. The Atari 800 (1979) and Commodore International's Commodore 64 (1982)

retained cartridge slots for console-style games, but were also capable home computers. Games designed for computers at first resembled arcade and video console titles, but early computer games took advantage of greater flexibility, inspired by complex paper-and-pencil role-playing games such as "Dungeons and Dragons," boardgames, and Crowther's "Adventure." The original Adventure linked Crowther's experiences as an explorer of Kentucky's Mammoth and Flint Ridge cave systems to the Tolkien-inspired fantasy world of role-playing games; written in FORTRAN for the PDP-10 computer, Adventure became the prototype for "interactive fiction," games featuring scripted story lines revealed as players typed responses to textual information provided by software. The numerous text-only adventures published by Infocom during the 1980s pushed the "adventure" genre further, beginning with the wildly popular "Zork" series. Other games such as the "King's Quest" series by Sierra On-Line (1983), military simulations and role-playing games published by Strategic Simulations Incorporated (founded in 1979), Richard Garriott's "Akalabeth/Ultima" series (1979), and the sports and multimedia titles of Electronic Arts (founded in 1982) extended the simulation and storytelling capacity of computer-based games. MUD (multi user dungeon), developed by Roy Trubshaw and C. Richard Bartle at the Univesity of Essex in 1978, combined interactive fiction, role-playing, programming and dial-up modem access to a shared computer to build a virtual world on the basis of social interaction as much as structured game play; hundreds of themed multiplayer MUDs, and BBS-based games were written during the 1980s and early 1990s.

In the late 1980s, a new generation of video consoles led by the Nintendo Entertainment System (1985) and the Sega Genesis (1989) offered improved graphics and also introduced battery-powered storage cartridges that enabled players to save games in progress. Games such as Shigeru Miyamoto's "Super Mario Brothers" (1985) and "The Legend of Zelda" (1987) for Nintendo or Square's "Final Fantasy" series (1987) took advantage of these capabilities to provide deeper game experiences, flexible character building and complex, interactive environments, encouraging comparisons between video games and other narrative media such as cinema. In the 1990s, computer game designers exploited the three-dimensional graphics, faster microprocessors, networking, hand-held and wireless game devices, and the Internet to develop new genres for video consoles, personal computers, and networked environments. The most important examples included first-person "shooters"—action games in which the environment is seen from the players view—such as id Software's "Wolfenstein 3-D" (1992), "Doom" (1993) and "Quake" (1996); sports games such as Electronic Arts' "Madden Football" (1989) based on motion capture systems and artificial intelligence; and massively multi-player games such as "Ultima Online" (1997) and "Everquest" (1998), combining traits of MUDs and graphical role-playing games to allow thousands of subscribers to create avatars and explore "persistent" virtual worlds.

See also **Entertainment in the Home; Technology, Arts and Entertainment; Technology and Leisure**

HENRY LOWOOD

Further Reading

Cassell, J. and Jenkins, H., Eds. *From Barbie to Mortal Kombat: Gender and Computer Games*. MIT Press, Cambridge, MA, 1998.

Herman, L. *Phoenix: The Fall and Rise of Videogames*, 3rd edn. Rolenta Press, Springfield, NJ, 2001.

Herz, J. C. *Joystick Nation: How Videogames Ate our Quarters, Won Our Hearts, and Rewired our Minds*. Little, Brown & Company, Boston, 1997.

Kent, S.L. *The Ultimate History of Video Games: From Pong to Pokémon and Beyond: The Story behind the Craze That Touched Our Lives and Changed the World*. Prima, Roseville, CA, 2001.

Loftus, G.R. and Loftus, E.F. *Mind at Play: The Psychology of Video Games*. Basic Books, New York, 1983.

Poole, S. *Trigger Happy: Videogames and the Entertainment Revolution*. Arcade Publishing, New York, 2000.

Sheff, D. *Game Over: How Nintendo Conquered the World*, 2nd edn. Gamepress, Wilton, CT, 1999.

Sudnow, D. *Pilgrim in the Micro-World: Eye, Mind, and the Essence of Video Skill*. Warner Books, New York, 1983.

Wolf, M.J.P., Ed. *The Medium of the Video Game*. University of Texas Press, Austin, 2001.

Computer Displays

The display is an essential part of any general-purpose computer. Its function is to act as an output device to communicate data to humans using the highest bandwidth input system that humans possess—the eyes. Much of the development of computer displays has been about trying to get closer to the limits of human visual perception in terms of color and spatial resolution.

The earliest output devices for early digital computers were lamp displays (Z3 by Konrad Zuse, 1941). An early digital computer, the EDSAC (electronic delay storage automatic calculator) developed at Cambridge University in 1949,

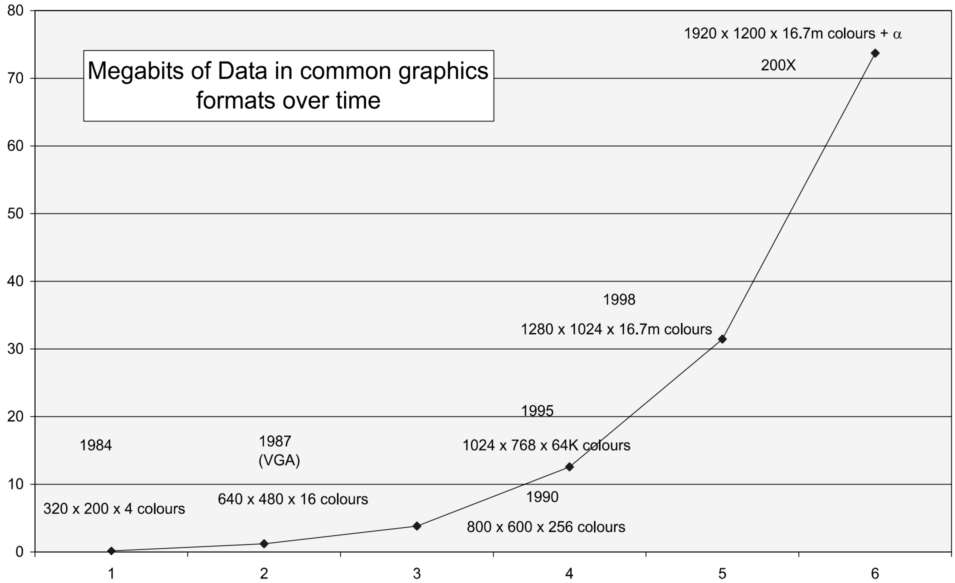


Figure 10. Increase in computer graphics bandwidth since the 1980s.

had a cathode-ray tube display monitor which could be used by programmers to view (as binary data with bright spots representing 1 and dim spots representing 0) the content of one memory store at a time. In the EDSAC there were 32 memory stores, or delay lines in total, each of which stored 32 words of 18 bits. The display monitors, with their flashing grid of dots, were only used to observe the progress of a program and monitor it. More usefully, numerical values were output to punched paper tape, or later to an attached teleprinter.

> "Input systems involving modified teletype equipment for inscribing keyboarded information were often integrated with teletypes and printers for the output device. This form of input and output continued to be used on some mainframes until the 1980s, but from the 1960s VDUs (visual display units) based on the cathode-ray tube replaced teletype as the output."

The development of computer graphics technology based on cathode-ray tubes in the early 1960s by Dr David Evans and Dr Ivan Sutherland at the University of Utah, led eventually to an explosion in computer display use. By the end of the century, around 150 million computer displays were being manufactured each year. These devices had an impact on the daily life of almost everybody in the developed world.

Most computer displays of the late twentieth century are based on cathode-ray tubes (CRTs), invented by Karl Ferdinand Braun in February 1897 in Aachen, Germany. In fact the name predates the identification of the electron as "cathode rays."

In the 1970s, when interactive terminals started to be needed in high volumes, the CRT was a natural choice as it was already in high-volume production because of the existing market for television use. The high volume of CRTs and the development investment to enable color TV gave the technology an overwhelming cost advantage over competing technologies.

The first computer displays used "vector scanning"; that is, the beam of the CRT was deflected using varying voltages to paint vectors, or lines, on the screen. Thus, when the computer wanted to paint a circle, the beam moved in a circular motion. Later displays used a fixed scanning system, called raster graphics and used in televi-

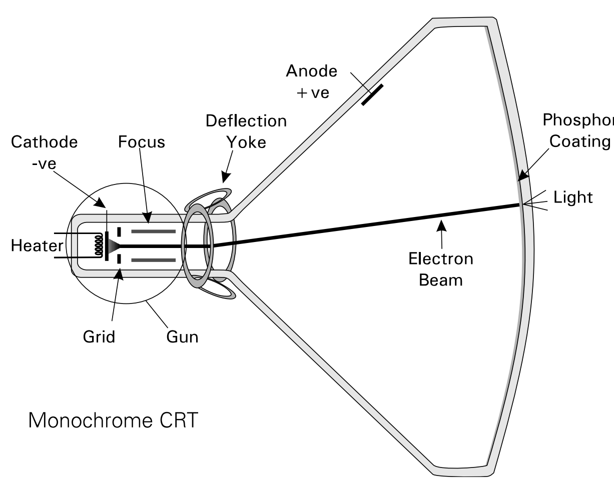


Figure 11. Line diagram showing operation of a CRT.

sions, where the image was stored in the memory of the computer as dots, or "pixels" (short for picture cells) that represent values for the color of the image at different screen locations. The computer scans through the memory and sends the values to the screen for display.

Cheaper and faster memory technologies enabled the widespread use of raster graphics in the 1970s and 1980s; CRT makers started to develop CRTs specially optimized for computer display use.

Mainframe and minicomputers used "terminals" to display the output. These were fed data from the host computer and processed the data to create screen images using a graphics processor. The display was typically integrated with a keyboard system and some communication hardware as a terminal or video display unit (VDU) following the basic model used for teletypes. Personal computers (PCs) in the late 1970s and early 1980s changed this model by integrating the graphics controller into the computer chassis itself.

Early PC displays typically displayed only monochrome text and communicated in character codes such as ASCII. Line-scanning frequencies were typically from 15 to 20 kilohertz—similar to television. CRT displays rapidly developed after the introduction of video graphics array (VGA) technology (640 by 480 pixels in16 colors) in the mid-1980s and scan frequencies rose to 60 kilohertz or more for mainstream displays; 100 kilohertz or more for high-end displays. These displays were capable of displaying formats up to 2048 by 1536 pixels with high color depths. Because the human eye is very quick to respond to visual stimulation, developments in display technology have tended to track the development of semiconductor technology that allows the rapid manipulation of the stored image.

CRTs have major disadvantages in bulk, weight, and power consumption and from the early 1970s onwards, there were many attempts to replace them with flat panel displays.

Liquid crystal displays (LCDs) were the main competitive technology. Early LCDs, developed in the late 1960s, had difficulty matching the response speeds, viewing angles, and color performance of CRTs and were much more expensive to manufacture. LCDs using an active matrix made an initial impact in mobile computers, where the lower bulk and power advantages made the cost premium acceptable. In the late 1990s, LCD makers developed active matrix TFT LCDs that could start to get close to CRTs for performance and price and adoption was then rapid even on the desktop. In the first decade of the twenty-first century, the CRT will be replaced by the LCD.

Most computers are used by individuals or very small groups at short viewing distances. However some are used to display to groups and two basic approaches are used—projection and direct view.

Early projection displays were based on high-brightness CRTs, but these were expensive and difficult to maintain and align. LCDs became the most popular imaging device for computer projectors, although the digital micromirror device (DMD) invented by Texas Instruments became popular for very portable projectors. In these projectors, small imaging devices (typically 23 millimeters diagonal) are used for creating an image that is illuminated by a very bright light and projected onto a screen using an optical system.

The main large-screen direct view display is the plasma display panel (PDP). Although originally conceived as a general-purpose desktop display, it proved difficult to make small pixels with efficient light output, so PDPs were developed for applications from 640 millimeters diagonal and above, designed for viewing from several meters or more. From the early development of monochrome PDPs in the 1970s and 1980s, by the 1990s full color PDPs up to 1.5 meters diagonal were in production.

For outdoor and public displays of very large size, such as those used in sports stadia, displays based on LED technology were developed. As has often been the case in display technology, blue devices proved difficult to discover and make and it was only in the 1990s that blue LEDs were first made, although red LEDs had been in production since the early 1960s and green shortly afterward. It has proved difficult to make LEDs with matched brightness, therefore LED displays are expensive, but can have a long life and high brightness.

Over many years there were attempts to develop "near to eye" personal displays, but problems in designing the high-quality low-weight and -cost optical systems and concerns about potential effects on health limited the use of this kind of display to professional and military applications.

See also **Computer-User Interface; Light Emitting Diodes; Liquid Crystals; Printers; Radar Displays**

ROBERT RAIKES

Further Reading

Foley, J.D., van Dam, A., Feiner, S.K., Hughes, J.F. *Computer Graphics, Principles & Practice*. 2nd edn. Addison-Wesley, 1995.

Keller, P.A. *The Cathode Ray Tube, Technology, History and Applications*. Palisades Press, New York, 1991.

Myers, B. *Display Interfaces, Fundamentals and Standards*. Wiley, Chichester, 2003.
Peddie, J. *High-Resolution Graphics Display Systems*. Windcrest/McGraw-Hill, 1994.

Computer Memory, Early

Mechanisms to store information were present in early mechanical calculating machines, going back to Charles Babbage's analytical engine proposed in the 1830s. It introduced the concept of the "store" and, if ever built, would have held 1000 numbers of up to 50 decimal digits. However the move toward base-2 or binary computing in the 1930s brought about a new paradigm in technology—the digital computer, whose most elementary component was an on–off switch. Information on a digital system is represented using a combination of on and off signals, stored as binary digits (shortened to bits): zeros and ones. Text characters, symbols, or numerical values can all be coded as bits, so that information stored in digital memory is just zeros and ones, regardless of the storage medium. The history of computer memory is closely linked to the history of computers but a distinction should be made between primary (or main) and secondary memory. Computers only need operate on one segment of data at a time, and with memory being a scarce resource, the rest of the data set could be stored in less expensive and more abundant secondary memory. The focus of this entry will be on primary memory technology, with attention to the early developments and the technological hallmarks.

A method of storing data on rotating "drums"—an electromechanical device—had emerged in the late 1930s. John Vincent Atanasoff successfully manufactured a device consisting of a rotating drum in which 1600 capacitors were placed in 32 rows. Later, magnetic coatings were used on drums and during World War II, this storage emerged as a reliable, rugged, inexpensive but slow memory device. Engineering Research Associated (ERA) produced a commercial device in 1947 that could store 65,000 32-bit words (over 2 million bits). Magnetic drum storage was to survive in later computers for secondary data storage, and later in disk platter form of continually shrinking proportions. It was the predecessor to the hard disk drive, the ubiquitous component of all later computers.

Early digital computers made use of electromechanical relays for binary logic. Early relay computers included Howard Aiken's automatic sequence-controlled calculator. Developed at Harvard University with the assistance of a group of International Business Machine (IBM) engineers, the Mark I was completed in 1944 and could store 72 numbers mechanically using electromagnetic decimal storage wheels, with the programming and data sequencing instructions fed into the machine via four punched tape readers. However the bulk, power requirements, and the delay time in the operation of relays were performance-limiting factors.

Thermionic valve (or vacuum tube) technology matured during World War II. Valves could operate as simple switches and a switching speed increase in the order of 1000 times over relays was realized. Although there were earlier prototypes, the completion of the ENIAC (electronic numerical integrator and computer) heralded the era of electronic computers. Built for the U.S. Army by the Moore School of Electrical Engineering at the University of Pennsylvania, it was only completed in late 1945 after the war. It consisted of 18,000 valves in which its core logic was implemented, as well as 1,500 relays which were used to store initial data loaded from a punched tape reader. Twenty "Accumulators" to store binary numbers during calculations were implemented with valve logic. Subsequently, purpose-built binary data storage valve technology emerged, although not without difficulties in production (see Selectron valve below).

The first generation of computers such as the Mark I and the ENIAC were purpose-built, and had the sequence of instructions "hard coded" or preprogrammed, for example with paper tape input. The British Colossus computer was another special purpose machine using electronic valves that implemented algorithms developed by Alan Turing and colleagues to crack encrypted German U-boat radio communications during World War II. However, mathematician John von Neumann approached the issue of storing programming code at a more fundamental logical level and initiated the concept of the stored program, whereby program code could be stored in memory much like the data was, meaning the setup for a new calculation could be expedited without rewiring. The EDVAC computer (electronic discrete variable automatic computer), first described in a draft report in June 1945 by von Neumann who was a regular visitor to the U.S. Army project at the Moore School, University of Pennsylvania, was the first design to implement the stored program concept—what came to be known as the "von Neumann" architecture. EDVAC was, however, just an idea at that stage, and not constructed by

the Moore Laboratory until 1949. Von Neumann and EDVAC's designers were keen to use purely electronic memory for stored program, not drums or relays.

From around the period of the mid- to late 1940s, emerging digital computers experienced what historian Thomas Hughes describes as a "reverse salient." This was a result of the scarcity of capacious and reliable memory technology and held back the advance of digital computers. Several electronic innovations in response to this were adaptations from technology developed during the war.

The ultrasonic delay line was developed for storing analog radar information, and was easily adapted to digital storage. The delay line worked by using a piezoelectric quartz crystal to convert an electrical signal into a sonic wave pulse, which then traveled through a liquid medium at the speed of sound in a long tube, and was then converted back to an electric signal. The difference in the speed of propagation meant that a number of bits of data could be stored. The delay line was a serial device—once a pulse had entered the tube there was no way to access it until it emerged at the other end. All operations were considered as serial transfers of numbers.

Delay lines were first built at the Bell Telephone Labs by William Shockley, and later at the Moore School for the Radiation Laboratory at the Massachusetts Institute of Technology (MIT) in 1943. Mercury acoustic delay lines were adapted for use with computers by J. Presper Eckert and John Mauchly in 1946 and selected for the EDVAC, which was actually not completed until 1952. Mercury delay lines realized a hundred times storage density saving over valve technology. The first use in digital computers was in Maurice Wilkes's EDSAC (electronic delay storage automatic computer) completed in 1949 at the Cavendish Laboratory in Cambridge, U.K., where 32 "tanks" of delay lines each stored 32 words of 18 bits (totaling about 2 kilobytes).

In response to the high cost and slowness of delay lines, Frederick Williams and colleagues at Manchester University in the U.K. developed the technique of using electrostatic cathode-ray display tubes as digital stores. The cathode-ray tube was another wartime development and the persistence of the phosphor display screen once illuminated by the cathode ray provided the memory. By 1948, a storage of 1024 bits was successfully implemented. Williams' colleague Tom Kilburn made improvements that increased the capacity to 2048 bits. The Williams–Kilburn tubes (commonly known as Williams tubes) were used on several of the early stored program computers, including the Manchester "Baby" (1948) and the Manchester Mark 1 which became operational in 1949, and the Institute of Advanced Study (IAS) machine spearheaded by von Neumann at Princeton, finally completed in 1951. The Williams tube memory had a big advantage over delay line memory in that it allowed fast random access (any memory location could be addressed and read directly). The Manchester Mark I was the first to store both its programs and data in RAM (random-access memory), as modern computers do.

In 1946 Vladimir K. Zworykin, Jan Rajchman and colleagues at the Radio Corporation of America (RCA) labs built the Selectron valve (or selective storage electrostatic tube) that could store 256 binary digits (bits) of information, and was described by Herman Goldstine (who directed the outpost of the U.S. Army's Ballistics Research Laboratory at the Moore School from 1942) as "a work of great engineering virtuosity." Like the Williams tube, the Selectron was also a random-access storage device. The Selector design however took over three years to reach the market, and was scaled down from the initial specification of 4096 bits storage. The revised Selectron was used in the 1953 JOHNNIAC at the Rand Corporation. The JOHNNIAC was one of the many machines based on the Princeton Institute of Advanced Studies (IAS) report authored in 1946 by John von Neumann, Arthur Burks, and Goldstine, based on their work on digital computers at Princeton following their involvement at the Moore School.

Magnetic core memory was first suggested by Jay Forrester of the Servomechanisms Laboratory at MIT and Andre Booth of the University of London, and was an important step in miniaturization and speed of computer memory. Small ferrite rings that were laid out in a two-dimensional matrix, with the magnetic property of hysteresis "remembering" an electric pulse. Each element of data was independently accessible via a row and column addressing system. Forrester and colleagues developed a successful prototype in 1949, and core arrays storing 1024 bits were used in their Whirlwind computer installed for the Navy's SAGE anti-aircraft defense system in 1958. Core memory became the dominant means of primary storage and was widely used on large commercial and scientific computers such as the IBM 701 in 1955, which was a significant commercial success and led to IBM's establishment as the dominant player in large systems. The IBM 7094 in 1963 had a core memory with over a million bits.

Ultimately it was the solid-state semiconductor that was to provide the memory technology of the computer revolution. The transistor was invented in 1947 by William Shockley at the Bell Laboratories after eight years of research on semiconductors for radar use. It was small and had very low power requirements and the switching speed was very quick compared to valves and core memory. Although it was a functional "switch" replacement for the valve as discussed above, it did not appear as computer memory until the late 1950s. There was significant momentum with the success of magnetic core memory in the industry, and the reliability of initial mass produced transistors was a contributing negative factor. Furthermore, Bell Labs was a regulated monopoly and limited by a court decision to the telecommunications business, although they did use transistors in special purpose computers built for the military use in intercontinental ballistic missiles around 1952. However, information about transistor technology was available to other manufacturers, with Philco successfully mass producing reliable and high-performance components. These were used in the SOLO, regarded as the first general purpose transistorized computer, completed by 1958. The commercialized version was called the TRANSAC which became available in 1960. Although the IBM 7090 (first installed in 1959) had a transistorized central processing unit, it still used magnetic core memory.

The development of integrated circuit (IC) technology began in 1959 with Jack Kilby of Texas Instruments. Collections of transistors were miniaturized onto a silicon "chip." It was only in the early 1970s that semiconductor memory was commercially used in computers after earlier pioneering use in purpose-built military and Apollo-era space guidance computers in the 1960s. The Super Nova produced by Data General in 1971 was the first commercial computer to use integrated circuit transistor-transistor logic (TTL) IC chip memory, after a less-successful research computer, the Illiac-IV, broke the tradition of use of magnetic cores by using a 256-bit memory chip made by Fairchild. IBM went down a proprietary route by using monolithic semiconductor memory in the System/370 Model 145 in 1971. Intel entered the IC memory market by producing a 1024-bit memory chip in 1970.

Memory chips were rapidly embraced by the emerging minicomputers, resulting in both compact size and more importantly, low cost. Later in the 1970s, these IC memory chips together with microprocessors provided the basis of the personal computer. With the advent of large-scale integration (LSI) and very large scale integration (VLSI), the cycle of increasing density of semiconductor memory fabrication continued unabated. Although the concepts and architecture of computers remained consistent with early designs, these techniques, competitive markets, and insatiable demand for personal computers globally, resulted in a doubling of memory density almost every year to the end of the century where up to a billion transistor elements on memory chips were commonplace.

See also **Computers, Early Digital; Computer Memory, Personal Computers; Integrated Circuits, Design and Use; Processors for Computers; Transistors; Vacuum Tubes/Valves**

BRUCE GILLESPIE

Further Reading

Ceruzzi, P.E. A history of modern computing. MIT Press, Cambridge, MA, 1998.
Eames, C. and Eames, R. *A Computer Perspective, Background to the Computer Age*. Harvard University Press, Cambridge, MA, 1973.
Goldstine, H.H. *The Computer: From Pascal to Von Neumann*. Princeton University Press, Princeton, 1972.
Hodges, A. *Alan Turing: The Enigma*. Burnett Books, 1983.
Wurster, C. *Computers: An Illustrated History*. Taschen, Koln, 2002.

Computer Memory, Personal Computers

During the second half of the twentieth century, the two primary methods used for the long-term storage of digital information were magnetic and optical recording. These methods were selected primarily on the basis of cost. Compared to core or transistorized random-access memory (RAM), storage costs for magnetic and optical media were several orders of magnitude cheaper per bit of information and were not volatile; that is, the information did not vanish when electrical power was turned off. However, access to information stored on magnetic and optical recorders was much slower compared to RAM memory. As a result, computer designers used a mix of both types of memory to accomplish computational tasks. Designers of magnetic and optical storage systems have sought meanwhile to increase the speed of access to stored information to increase the overall performance of computer systems, since most digital information is stored magnetically or optically for reasons of cost.

The first systems for the magnetic storage of large amounts of digital information were devel-

oped in the early 1950s. They were designed to replace the punch card record systems then in common use in large business and government organizations. These early magnetic systems were based on reel-to-reel tape recorder technology developed in the 1930s and 1940s for sound recording. In adapting these systems for computer use, the controls for tape movement were made fully electronic so that they could be controlled by the central processing unit (CPU) in the computer. The only other significant innovation was the addition of a system whereby a loop of tape was maintained using vacuum to buffer the rapid starts, stops, and changes in direction of the tape, considerably reducing the chance of tape breakage. It is this vacuum system that accounts for the common appearance of mainframe computer tape drive systems—the tape reels mounted in the upper part of a tall cabinet and the vacuum mechanism concealed behind the panel below.

While an inexpensive form of storage, tape drives were very slow due to the one-dimensional arrangement of information. Thus, by the 1960s they were replaced by disk drives for most computer applications. However, due to their very low cost per bit for storage, tape drives continued to be used for backup storage as information was regularly copied on hard disks and other storage methods. The form of these backup tape drive systems changed over time, mirroring changes in sound, and later videotape systems. In particular, tape drive systems moved from open reel to cassette systems by the 1980s. At the end of the twentieth century, tape systems continued to be used for backup copies.

The disk drive, the most commonly used computer mass storage device of the twentieth century, was invented at International Business Machines (IBM) in the early 1950s by a team led by Reynold B. Johnson and Louis Stevens, and first marketed in 1957. An outgrowth of earlier drum memory systems, the disk drive consists of one or more flat disks that rotate at high speed. Information is magnetically recorded on and read from the surface of the disk using an electromagnetic coil, the read/write head. Information is recorded in a series of concentric tracks, and the head is moved between tracks by an actuator. Thus, in contrast to the one-dimensional storage on a tape drive, the two-dimensional form of the disk drive allows for much faster access to a particular bit of information, albeit at increased cost and complexity.

The basic mechanical operation of disk drives has not changed since their initial development. However, the capacity for data and the speed of access to it has increased by many orders of magnitude. This has resulted from a large number of gradual improvements rather than a single breakthrough. Three major areas of improvement account for most of the speed increase. First, the development of improved magnetic alloys allowed for greater recording densities without loss of accuracy. Second, improved manufacturing methods produced smoother disk surfaces. As a result, the recording head could operate closer to the disk surface, making narrower tracks and so increasing density. Finally, better recording heads and more accurate actuating mechanisms to guide them also allowed for narrower recording tracks and greater density. As a result, the average diameter of disk drives has dropped from 12 inches (300 millimeters) in the 1960s to under 3 inches (76 millimeters) at the end of the twentieth century. The smaller size and greater density of the disks means that data access becomes faster since heads do not have to move as far to read information.

The operation of removable magnetic media such as floppy disk drives, zip-drives, and the like is identical to that of hard disks. Removable media operate with lower precision than hard drives, and thus storage densities are lower. However, the advantage of removable media is its portability—data can be moved from one computer to another. In the late 1990s, a hybrid type of removable media emerged, a self-contained disk drive that could be easily plugged into computer systems without a separate power supply.

Optical storage systems operate in a fashion similar to disk drives. However, information is recorded using a laser to mark the surface of the disk, and the resulting marks are scanned using the laser and a photoelectric sensor. The primary advantage of optical storage is its resistance to bulk erasure. Unlike magnetic storage, which can be erased by exposure to a strong magnetic field, optical storage is more stable. Although in theory optical systems can have recording densities similar to magnetic media, in practice during the twentieth century magnetic systems had higher densities due to engineering differences.

The most common computer-related optical storage medium in the late twentieth century was the compact disk (CD), based on the same CD developed for recorded music in the 1970s. Information is recorded on the disk in one continuous spiral rather than in a concentric ring. The introduction of new recording media in the 1990s allowed the CD, initially a write-once, read-many system, to function much like the disk drive. Due to low cost and stability, the CD became the

media of choice in the late 1990s for the distribution of software, an application previously dominated by removable floppy disks. However, magnetic recording continued to dominate hard disk applications due to lower cost and higher densities.

Both magnetic and optical recording methods have upper physical limits on recording density, but by the end of the twentieth century those limits had not yet been reached. Due to their low cost per bit, it is likely that these types of recording will be used in computer systems for some time to come.

See also **Alloys, Magnetic; Audio Recording, Compact Disk; Computer Memory, Early**

MARK CLARK

Further Reading

Bell, A.E. Next generation compact discs. *Scientific American*, 275, 1, 42–46, 1996.
Daniel, E.D. Mee, C.D. and Clark, M.H. *Magnetic Recording: The First 100 Years*. IEEE Press, 1999.
Camras, M. *Magnetic Recording Handbook*. Van Nostrand Reinhold, New York, 1988.

Computer Modeling

Computer simulation models have transformed the natural, engineering, and social sciences, becoming crucial tools for disciplines as diverse as ecology, epidemiology, economics, urban planning, aerospace engineering, meteorology, and military operations. Computer models help researchers study systems of extreme complexity, predict the behavior of natural phenomena, and examine the effects of human interventions in natural processes. Engineers use models to design everything from jets and nuclear-waste repositories to diapers and golf clubs. Models enable astrophysicists to simulate supernovas, biochemists to replicate protein folding, geologists to predict volcanic eruptions, and physiologists to identify populations at risk of lead poisoning. Clearly, computer models provide a powerful means of solving problems, both theoretical and applied.

Scientific models, or representations of nature, date to antiquity, when astronomers developed mathematical methods to predict planetary movements. However, computer models only originated in the mid-1940s. Indeed, the history of computer models is intertwined with that of computers. The second digital computer, the ENIAC (electronic numerical integrator and computer), was used to simulate an early design of the hydrogen bomb. Before the ENIAC, physicists had to use slide rules to calculate the power of nuclear explosions. The invention of high-speed digital computers made it possible to speed up the calculation process and thereby perform faster simulation experiments by programming mathematical equations representing the laws of nature. Computers obtain solutions to differential equations by evaluating variables at specific points at given time intervals. In this way, a computer can simulate the development of a thunderstorm by evaluating pressure, temperature, wind velocity, humidity, and other variables at thousands of points in space and time. Models are thus mathematical representations of physical phenomena.

Although computing power has increased exponentially over the past six decades, modelers must still make trade-offs between resolution and complexity. A modeler can increase the accuracy and realism of a simulation by adding more spatial points and shortening the time intervals, or by adding more laws of nature to the set of equations. However, while higher-resolution models decrease the levels of uncertainty associated with computer modeling, such actions increase computing time and power, demands on personnel, and funding obligations.

Because it remains difficult to integrate all the variables of complex systems, even the most sophisticated computer models contain shortcomings. Such limitations include reliance on simplifying assumptions, spatial resolution problems, difficulties in validating models with observational data and experimental results, problems in representing small-scale processes in terms of large-scale variables, and weak agreement between different models. Such deficiencies have provoked intense debates regarding the scientific standards of model trustworthiness, especially the accuracy of modelers' assumptions. These debates are not limited to the scientific community, since models have explicit policy-making value. Computer models are therefore interesting not only for the ways in which they have changed the practice of science, but also for the political and epistemological questions raised by their use.

The significant political role of computer models is exemplified by the ongoing international debates over global climate change and the greenhouse effect. Today's global climate models evolved from weather-prediction models and general circulation models of the 1950s and 1960s. During the 1970s and 1980s, larger, faster, cheaper computers facilitated the construction of models capable of analyzing both atmospheric and oceanic general circulation. Such "coupled" models allow scientists

to study the effects of increasing carbon dioxide levels on the global climate, and other phenomena that cannot be subjected to controlled laboratory tests, including acid rain, ozone depletion, and nuclear winter. Politicians have used data generated by such models to develop controversial policies, such as the 1997 Kyoto Protocol, a treaty designed to slow global warming by requiring 38 developed nations to reduce greenhouse gas emissions to an average of 5.2 percent below 1990 levels by 2012.

Another important political role played by computer modeling relates to national security. Following the end of the Cold War and the U.S. government's 1992 moratorium on nuclear testing, U.S. officials seeking to build domestic support for the Comprehensive Nuclear Test Ban Treaty while keeping the national nuclear weapons laboratories open, developed the Stockpile Stewardship Program and Accelerated Strategic Computing Initiative (ASCI). As a result, supercomputer models have been used since the mid-1990s to maintain the U.S. nuclear stockpile without detonating actual bombs. The political need to simulate the incredibly complex physics of nuclear explosions has led to tremendous advances in computational power. Over several months beginning in 2001, researchers at the Lawrence Livermore and Los Alamos national laboratories used the world's fastest supercomputer, IBM's ASCI White—capable of performing 12.3 trillion calculations per second (teraflops)—to conduct the first complete three-dimensional simulation of a thermonuclear explosion. ASCI officials anticipate the creation of 100-teraflop supercomputers, a technological advance with applications far beyond stockpile stewardship. Indeed, critics have long argued that such efforts to scale up supercomputing will facilitate the design and development of new nuclear weapons, not just the maintenance of ageing armaments.

Despite the awesome advances in supercomputer modeling techniques brought about by weaponry testing and weather forecasting, models still cannot represent complex real-world processes with absolute precision. For this reason, some groups oppose the use of model-generated data in the policymaking arena. Most notably, global warming skeptics point to the inability of global climate models to predict regional changes, and scorn modelers' assumptions and simplifications as "internal fudge factors" which generate "garbage-in, garbage-out" simulations. Such charges have led to antagonistic debates on the relationship between model trustworthiness and credible science, as well as initiatives to improve computer modeling uncertainty analysis.

The political use of computerized models has also stimulated discussions of the philosophical and epistemological implications of computer modeling. During the 1990s, some scientists and scholars in the field of science studies asserted that computer models cannot be validated, since models function as inductive arguments, which according to the philosopher Karl Popper, can only be proven false, not true. One organization that relies on computer models, the United Nations-sponsored Intergovernmental Panel on Climate Change, responded by replacing the word "validation" with "evaluation" to convey the extent of correspondence between models and the real-world processes they represent. Other analysts questioned how policymakers can expect high levels of model-based certainty, since many political decisions are made under uncertain conditions. As with any powerful new technology, debates over the trustworthiness of computer modeling techniques will persist as long as unacceptable levels of uncertainty remain. Nevertheless, computer models will likely continue to transform both the practice of science and the production of policy-relevant knowledge.

See also **Computers, Uses and Consequences**

CHRISTINE KEINER

Further Reading

Cipra, B. Revealing uncertainties in computer models. *Science*, 287, 960–961, 2000.

Edwards, P.N. A brief history of atmospheric general circulation modelling, in *General Circulation Model Development: Past, Present, and Future*, Randall, D.A., Ed. Academic Press, San Diego, 2000.

Kaufmann III, W.J. and Smarr, L.L. *Supercomputing and the Transformation of Science*. Scientific American Library, New York, 1993.

Konikow, L.F. and Bredehoeft, J.D. Ground-water models cannot be validated. *Adv. in Water Resour.*, 15, 75–83, 1992.

Kwa, C. Modelling technologies of control. *Sci. Cult.*, 4, 20, 363–391, 1994.

Oreskes, N. Evaluation (not validation) of quantitative models. *Environ. Health Perspect.*, 106, 6, 1453–1460, 1998.

Oreskes, N. Why believe a computer? Models, measures, and meaning in the natural world, in *The Earth Around Us: Maintaining a Livable Planet*, Schneiderman, J.S., Ed. W.H. Freeman, New York, 2000.

Paine, C.E. A case against virtual nuclear testing. *Scientific American*, 281, 74–79, 1999.

Shackley, S., Risbey, J., Stone, P. and Wynne, B. Adjusting to policyexpectations in climate change modelling: an interdisciplinary study of flux adjustments in coupled atmosphere-ocean general circulation models. *Climatic Change*, 43, 413–454, 1999.

Weiss, P. Computer simulates full nuclear blast. *Science News*, 161 189, 2002.

Computer Networks

Computers and computer networks have changed the way we do almost everything—the way we teach, learn, do research, access or share information, communicate with each other, and even the way we entertain ourselves. A computer network, in simple terms, consists of two or more computing devices (often called nodes) interconnected by means of some medium capable of transmitting data that allows the computers to communicate with each other in order to provide a variety of services to users.

Using the above definition, a typical computer network must include the following:

- A set of computing nodes which may range from simple personal digital assistants for retrieving electronic mail (e-mail) all the way to the largest supercomputers.
- A data transmission medium over which the nodes can communicate with each other. This medium can be optical fiber, coaxial cable, copper twisted pairs, or even air and free space for wireless transmission of data. In effect, the transmission media can be considered as the roads and highways over which data and information are transmitted.
- An interface (often called a NIC or network interface card) between the computing node and the data transmission medium to transform the digital signals used in the computer into a form suitable for the transmission medium (for example, light signals for optical fiber or electromagnetic waves for wireless transmission).
- A communication or network protocol which is the set of rules and conventions that govern the effective transmission of data over the network to ensure that communicating nodes ‘‘speak the same language,’’ recognize, and understand each other.
- variety of devices (often called network electronics) used to perform a variety of network control tasks such as traffic switching, filtering, aggregating, routing, protocol translation, and signal conversion from one transmission medium to another (for example, from an optical fiber medium between buildings to a twisted copper pair medium within buildings).

Historically, the first communications networks were telegraphic. Communication protocols were developed to relay messages from one network station to another (‘‘message switching’’). The expansion of telephone networks in the late nineteenth century meant that each telephone could not be connected directly to every other telephone; automatic switching was needed (‘‘circuit switching’’). By the 1960s when the first computer networks were being developed, similar switching for data communications over telephone lines was needed. (The development of the first resource-sharing computer networks (at the National Physical Laboratory in England and ARPAnet) and the concurrent development of ‘‘packet switching’’ are described in the entries on Electronic Communications, Internet, and Packet Switching.)

A given computer network can be characterized by a number of attributes, with the most important being geographic size, protocol and type of access control used, and the way it is funded. Networks that span a local area such as one or more neighboring buildings are called local area networks, or LANs. Ethernet is by far the most widely used LAN technology today. It was developed in the mid-1970s at the Xerox Palo Alto Research Center to interconnect a number of desktop computers and printers. It was so successful that it quickly became a standard. Since those days, there have been numerous enhancements and there are standards defining Ethernets that support different network topologies and a variety of transmission media at rates of 10,100 and 1,000 million bits per second.

Networks distributed over a wide area (distances of tens to thousands of kilometers) are called wide area networks or WANs, while metropolitan area networks (MANs) span distances between those of a LAN and a WAN. The transmission media for LANs is usually provided by the organization owning the LAN, while the transmission media for WAN connectivity are usually leased from telephone or common carrier companies.

Computer networks can also be characterized by the way they are funded. Private networks, for example, are usually owned by some entity that confines use of the network and its services to its employees or customers. Public networks are owned by entities that offer network services to any individual or organization that wishes to subscribe and pay for the services provided. Networks that provide e-mail and Internet access services to any user on a fee-for-service basis are examples of public networks. Finally, community networks refer to networks that are managed and

supported by a local community and provide services specially tailored to its community of users.

Client/server and peer-to-peer are phrases used to describe networks with different types of hierarchy and access control. In a client/server network, there is at least one computer that is dedicated as a server to some kind of service (for example e-mail or Web service) which controls access from other computers that are its clients. A peer-to-peer network, on the other hand, does not have any dedicated servers or hierarchy among its computers. Instead, users must make decisions regarding who gets access to what, which adversely impacts system security and consequently such networks are limited to small numbers of computers.

In the very early days, computer networks were relatively small, and computer or modem vendors tended to develop proprietary protocols that allowed their machines to communicate with each other, but not necessarily with those made by others. This created the need to interconnect two or more compatible or incompatible networks together to create an Internet, or a network of networks. In the mid 1970s, TCP/IP (transmission control protocol/internet protocol) was developed and became the main communication protocol using the Unix operating system and its various derivatives. Today, TCP/IP is the de facto standard and is available for almost all computers. When sending a file from one machine to another, TCP breaks it into a number of appropriately sized data packets, which also include other information such as the type of packet, source, and destination addresses. IP then takes over, routing those packets from the source to the destination, where TCP takes over and reassembles them into the original file. Together, these two protocols allow data and control information to be "packetized," addressed, and routed to the destination where it is reassembled and appropriately processed.

Today, the great majority of computer networks are connected to form the global Internet, which uses TCP/IP. When a user wants to access a Web server such as www.xyz.com, the symbolic destination domain, xyz.com, is translated into a specific IP address, and the user is connected to that server. Once there, the browser on the client (user) machine allows the user to look at specific files of various kinds. Similarly, when an e-mail is sent to userid@xyz.org, the source mail server sends the message to the destination server xyz.org which then stores the file in the mail folder of userid, until the latter logs into xyz.org, reads the e-mail, and deletes it.

Figure 12 shows two LANs where computers are connected via switches and routers to the global Internet. Today, the latter interconnects millions of such LANs worldwide, with a total of tens of millions of computers providing a variety of services to hundreds of millions of users. Its total traffic has been growing exponentially, and it is expected that this growth will continue as new generations of Internet applications are developed, especially those that require the digitization and transmission of voice or video packets in addition to data.

See also **Electronic Communications; Internet; Packet Switching; Telephony, Digital; World Wide Web**

JOHN SOBOLEWSKI

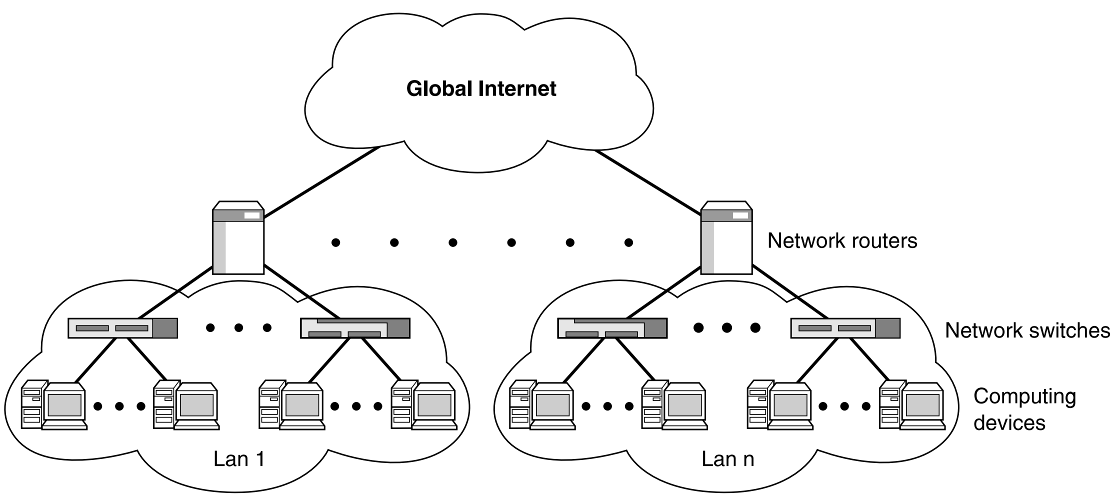


Figure 12. Diagram of a computer network showing how network switches interconnect computers within LANs, which are then connected via routers to the Internet.

Further Reading

Comer, D. and Stevens, D. *Internetworking with TCP/IP. Vol. I: Principles, Protocols and Architecture*. Prentice Hall, Englewood Cliffs, NJ, 1994. A description of principles of communication protocols with an emphasis on TCP/IP.

Metcalf, R. and Boggs, D. Ethernet: distributed packet switching for local computer networks. *ACM Comm.*, 19, 7, 395–403, 1976. The classic paper describing the first Ethernet.

Peterson, L. and Davie, G. *Computer Networks: A Systems Approach*. Morgan Kaufman, San Francisco, 2000. A recent book describing various aspects of computer networks.

Ralston, A. *et al.*, Eds. *Encyclopedia of Computer Science*. Nature Publishing Group, Oxford, and Grove's Dictionary, New York, 2000. This contains a number of articles describing various aspects of computing and computer networking.

Sobolewski, J. Data transmission media, in *Encyclopedia of Physical Science and Technology*, 3rd edn, vol. 4, Meyers, R., Ed. Academic Press, London, 2001. Describes various data transmission media in use today.

Computer Science

Computer science occupies a unique position among the scientific and technical disciplines. It revolves around a specific artifact—the electronic digital computer—that touches upon a broad and diverse set of fields in its design, operation, and application. As a result, computer science represents a synthesis and extension of many different areas of mathematics, science, engineering, and business.

The tide that lifted science and engineering in general following World War II lifted computing as well, albeit in a less straightforward fashion. It was the need to speed up the calculation of ballistic firing tables (beyond what was achievable using mechanical differential analyzers, essentially analog calculators) after all, that led to development of the ENIAC (electronic numerical integrator and computer), completed in 1945. Science and engineering emerged from World War II and entered the Cold War with a new sense of national importance attached to them. As a result, government funding of scientific and technical research and training, especially in universities and particularly in the U.S. and the U.K., surged.

However, the interdisciplinary nature of computing complicated development of a computing discipline both politically and substantively. Establishing the jurisdiction of a profession or discipline by definition involves political issues of control and legitimacy. Since computing did not obviously and by consensus fall within the purview of any single group of practitioners or academics, many laid claim to it and to the prerogatives and resources that accompanied it. Existing academic disciplines, including applied mathematics, electrical engineering, and physics (varying on a university-by-university basis), attempted to keep computing under their respective institutional wings. Industrial practitioners, meanwhile, took issue with academic approaches in general, demanding a discipline that bore some connection with what they actually did. In the U.S., for example, this meant that even as mathematics and electrical engineering departments on some campuses engaged in a tug of war to define computer science, the Association for Computing Machinery, with its academic bent, was engaged in a similar struggle with the Data Processing Management Association, with its industrial orientation.

Substantively, no one had a clear or dominant definition of what computing practitioners did or what knowledge constituted the core of a computing discipline. As a result, efforts to define it tended to reflect the disciplinary and professional backgrounds of the people doing the defining. Computer science (the very term prompted some heated arguments) ended up adopting an eclectic mixture of knowledge from a range of existing disciplines and from industry as well. From electrical engineering came such things as switching theory and circuit analysis and design. From applied mathematics came numerical methods and algebraic analysis. From linguistics came syntactic and semantic analysis. From psychology came cognitive theory. From business came information management. All of these, and others as well, were adapted and extended to focus on computing, leading to distinctive areas such as computer architecture (the organization of computer hardware), algorithmic analysis (the determination in principle of the time and storage efficiency of algorithms), computability theory (the analysis of which classes of problems are in principle solvable by a computer), artificial intelligence (the development of computer systems that can learn and otherwise replicate human intelligence), human–computer interaction (the analysis and application of the psychological processes governing communication between humans and computers), information storage and retrieval (the organization and use of databases), software engineering (the development of software in a disciplined and systematic fashion), and programming languages (the analysis and design of programming languages using formalized principles and notations), among others.

Computing professionals worried about how to unify this range of concerns in a fashion deeper than their common focus of the computer. This search for "fundamental principles" akin to those in established disciplines generated a great deal of angst, and to some extent still does. However, identifying principles with both the requisite breadth and the desired depth has proven very problematic. Often the two qualities represent a trade-off.

Computing professionals also sought to be viewed as such by others. Industrial practitioners sought this recognition from the both the government and the general public. In the U.S., for example, programmers and system analysts were outraged when, in 1972, the Department of Labor ruled that they were not professionals. At the same time, academics openly fretted about their professional standing relative to other academic fields and the potential impact on government research funding. While some of this same anxiety was evident in Europe, it was generally milder. This was partly due to the legitimacy lent by different professional structures, which generally awarded professional societies in Europe much more authority to regulate their professions. The British Computer Society, for example, has full and explicit authority to accredit curricula and to nominate Chartered Engineers. Moreover, debates about whether computer science was in fact a science were more muted in those parts of Europe that placed this collection of knowledge under the rubric of "informatics." (This represented a certain irony as the European inclination toward mathematical formalism actually meshed better with the "science" label than some of the American formulations.) Even today, computing professionals continue to express concern over their professional status as reflected in research funding and the perceptions of public and peers alike.

In hindsight, part of the difficulty of defining a computing discipline was the tendency of the individuals and institutions involved to treat the process as a zero-sum game. From this perspective, there could only be a single computing discipline and its orientation and content therefore were subject to a great deal of dispute. More recently, however, there has been growing acceptance of the notion that computer science is only one of several related but distinct computing disciplines including computer engineering (focused on computer hardware), software engineering (focused on the software that runs on the hardware), information systems (oriented toward business), and, in Europe, various specialized forms of informatics (focused on the application of computing within a specific domain such as medicine). Relatively specialized disciplines may possibly prove more amenable to the discovery or development of fundamental principles as well. The realization that computer science need not serve as the sole discipline for computing may ease a number of the stresses and strains that have characterized the development of computer science for most of its brief history.

See also **Artificial Intelligence; Software Engineering**

STUART S. SHAPIRO

Further Reading

Aspray, W. Was early entry a competitive advantage? US universities that entered computing in the 1940s. *IEEE Ann. Hist. Comput.*, 22, 3, 42–87, 2000. An analysis of the early institutional development of computer science at five major U.S. universities.

Croarken, M. The emergence of computing science research and teaching at Cambridge, 1936–1949. *IEEE Ann. Hist. Comput.*, 14, 4, 10–15, 1992. An analysis of the institutional origins of computer science at Cambridge University in the U.K.

Denning, P., Comer, D., Gries, D., Mulder, M., Tucker,A., Turner, A.J. and Young, P. Computing as a discipline. *ACM Comm.*, 32, 1, 9–23, 1989.

Ensmenger, N. The "question of professionalism" in the computer fields. *IEEE Ann. Hist. Comput.*, 23, 4, 56–74, 2001. An analysis of professionalization efforts in computing in the U.S. in the 1950s and 1960s.

Giordano, R. Institutional change and regeneration: a biography of the computer science department at the University of Manchester. *IEEE Ann. Hist. Comput.*, 15, 3, 55–62, 1993. An analysis of the institutional development of computer science at the University of Manchester in the U.K.

Hartmanis, J. and Lin, H., Eds. *Computing the Future: A Broader Agenda for Computer Science and Engineering*. National Academy Press, Washington D.C., 1992.

Report of the Association for Computing Machinery's Task Force on the Core of Computer Science, presenting an intellectual framework for the discipline of computing.

Report of the U.S. National Research Council's Committee to Assess the Scope and Direction of Computer Science and Engineering. Chapter 6: What is computer science and engineering?.

Shapiro, S. Boundaries and quandaries: establishing a professional context for IT. *Inf. Technol. People*, 7. 1. 48–68, 1994. An analysis of professionalization efforts in computing in the U.S. and the U.K. focusing primarily on the 1970s and 1980s.

Computer-Aided Design and Manufacture

Computer-aided design and manufacture, known by the acronym CAD/CAM, is a process for

manufacturing mechanical components, wherein computers are used to link the information needed in and produced by the design process to the information needed to control the machine tools that produce the parts. However, CAD/CAM actually constitutes two separate technologies that developed along similar, but unrelated, lines until they were combined in the 1970s.

Computer-Aided Design

Today, engineers use computers in the process of designing most products, from mechanical components to chemicals. However, computer-aided design is a relatively new practice. While engineering applications have been central to the development of the computer from its World War II origins, the commonplace use of computers in engineering design only dates to the introduction and diffusion of the minicomputer in the 1970s. Computer-aided design spread rapidly with the introduction of the microcomputer, especially the personal computer, or PC, in the 1980s.

As was the case with many of the technologies of the postwar period, research into computer-aided design began under the auspices of the military. The first electronic computers were designed and built for the war effort and used for ballistics calculations, cryptography, and atomic bomb calculations. The application of these new machines to an array of engineering problems followed shortly after the war. Finite-element analysis, the most successful method of computer-aided design, originated in airframe design in the early 1950s. In the 1940s, the introduction of the jet engine on military planes created new structural problems in the design of airframes. To take advantage of the faster speeds generated by jet engines, plane designers moved to a swept-back wing shape. They also moved to minimize the weight of the plane's structure. For structural engineers working on airframe analysis, the old methods of analysis were not accurate enough to allow them to optimize the airframe's design. At Boeing, analyzing the stiffness of an airplane's wing became the impetus to develop a new method of structural analysis. Ray W. Clough, a professor of civil engineering at the University of California, Berkeley who was working at Boeing in 1952 and 1953, traced the inaccuracy of the existing methods of analysis to the fact that the wing was modeled strictly as a frame, while the metal skin covering the frame was ignored. At the time, no method existed for mathematically modeling the metal skin. M. J. Turner, the head of Boeing's Structural Dynamics Unit, suggested that the team consider the patterns of strain within the skin as a way to take into account the stiffness of the skin. The wing's skin could be idealized as a mesh of arbitrary elements. Each element would add stiffness to the wing in a proportion drawn from the strain patterns determined experimentally. This mesh could be represented mathematically as a matrix, and the Boeing engineers knew that matrix notation lent itself to digital computation.

Coined finite-element analysis, or FEA, by Clough in 1960, the spread of the new method depended on access to high-speed digital computers. As a result, through the 1960s the use of the method was mainly confined to large corporations, usually with defense connections, though many companies in the automobile industry also came on board. As minicomputers became available and more affordable in the 1970s, FEA also became more accessible to smaller and mid-sized companies. As these computers offered greater graphical capabilities, engineers could literally draw their designs using computer-aided drafting programs, and feed those graphical models into FEA programs. With the introduction and remarkable spread of personal computers in the 1980s, FEA became accessible to every company, and its integration with software for graphical representation became complete. Commercial FEA packages were often bundled with drafting packages. FEA was the core of an information processing system for engineers. Using this type of software, engineers could perform computer-aided analysis of everything from structural analysis to heat transfer. Regardless of particular application, the key to the method remained the notion of laying a mesh over the design and dispersing it into arbitrary, finite elements that can be mathematically represented as matrices.

Computer-Aided Manufacturing

Like computer-aided design, the most successful development in computer-aided manufacture began as a collaborative project between a corporation and academe under the auspices of a military contract. While the notion of automating machine tools predated World War II, the direct development of tools to actually control the processes of machining began at the Massachusetts Institute of Technology (MIT) in the late 1940s. In 1949 John T. Parsons, of the Parsons Corporation, contacted the MIT Servo-Mechanisms Laboratory for advice on the mathematics connected with Parsons' contract to produce an automatic contour cutting mill for the U.S. Air Force. The Parsons Corporation

had made its name machining helicopter rotor blades, and was looking to get into the business of machining wing sections for new Air Force prototype airplanes. Parsons believed that by using punched cards to feed information to machine tools he could increase both the accuracy and speed of production. The partnership between Parsons and MIT coined the phrase numerical control, or NC, to describe the operation at hand. NC became the heart of computer-aided manufacture. With NC, each part to be machined could be described mathematically. The machine tool's motions were controlled by servomotors, which responded to digital commands programmed into the machine to determine a path for the cutting head to follow.

Parsons dropped out of the project in 1950 when he realized that fulfilling his Air Force contract did not entirely mesh with MIT's plan to fully automate machining using digital servo control. Still, he is rightly credited as the inventor of NC. Work on the automatic milling machine continued at MIT until 1956. Under MIT's direction, research into automation of machine tools focused more on control and feedback systems than on machining. Still, MIT's system of automation became commercially available in the early 1950s, through a Cambridge, company called Ultrasonic, with very close ties to the servo lab.

Like FEA, the diffusion of NC machining initially stayed close to the defense industries from which it was developed. Only with the development of cheaper, more powerful computers could NC become common outside the rarified atmosphere of military contracts.

CAD/CAM

In the 1970s, cheaper, more powerful minicomputers meant that many engineering companies were using the same computer to design and manufacture parts. The common use of computers led to the marriage of CAD and CAM, so that the information produced in the design of a component could be directly transferred through the computer to the shop floor. The graphical representation of components on the computers played a key role in putting CAD and CAM together. Engineers laid out a part on the computer in order to perform various analyses on it; that same information could then be used to set up an operation for producing the component. Again, the Air Force played a central role as a powerful proponent of CAD/CAM with its $100 million integrated computer–integrated manufacturing, or ICAM program. Through ICAM, the Air Force was able to foster corporate and academic, leading to faster, less risky development and dispersal of new CAD/CAM technology.

At the turn of the twenty-first century, parts as varied as airplane wings and fishing rods are designed and manufactured using desktop computers. CAD/CAM stands for a different way of making artifacts, as well as an integration of conceptual design and shop floor production. It has also ushered in a new method of oversight, whereby one individual can supervise and control both design engineering and manufacture on a single computer.

Ann Johnson

Further Reading

Bennett, S. *A History of Control Engineering, 1930–1955*. Institute of Electrical Engineers, London, 1993.

Diebold, J., *Automation*. American Management Association, New York, 1983.

Downey, G. CAD/CAM saves the nation?: Toward an anthropology of technology. *Knowledge Soc.*, 9, 143–168, 1992.

Johnson, A. From Berkeley to Boeing: civil engineers, the Cold War, and the origins of finite element analysis, in *Growing Explanations: Historical Perspectives on the Sciences of Complexity*, Wise, M.N., Ed. Duke University Press, Durham (in press).

Mindell, D.A. *Between Human and Machine: Feedback, Control, and Computing Before Cybernetics*. Johns Hopkins University Press, Baltimore, 2002.

Noble, D.F. *Forces of Production: A Social History of Industrial Automation*. Oxford University Press, Oxford, 1984.

Computers, Analog

Paralleling the split between analog and digital computers, in the 1950s the term analog computer was *a posteriori* projected onto pre-existing classes of mechanical, electrical, and electromechanical computing artifacts, subsuming them under the same category. The concept of analog, like the technical demarcation between analog and digital computer, was absent from the vocabulary of those classifying artifacts for the 1914 Edinburgh Exhibition, the first world's fair emphasizing computing technology, and this leaves us with an invaluable index of the impressive number of classes of computing artifacts amassed during the few centuries of capitalist modernity. True, from the debate between "smooth" and "lumpy" artificial lines of computing (1910s) to the differentiation between "continuous" and "cyclic" computers (1940s), the subsequent analog–digital split became

possible by the multitudinous accumulation of attempts to decontextualize the computer from its socio-historical use alternately to define the ideal computer technically. The fact is, however, that influential classifications of computing technology from the previous decades never provided an encompassing demarcation compared to the analog–digital distinction used since the 1950s.

The list of what is currently placed beneath the heading of pre-electronic analog computers seems inexhaustible: calendars; sundials and orreries; astrolabes and planetariums; nomograms and graphs; special and general purpose slide rules; slide rule-based assemblages; computing boards and tables; material models of all kinds (including scale models, linkages, artifacts with integrators and differentiators, curve tracers and kinematic mechanisms of which the most known are the various planimeter configurations); harmonic analyzers and synthesizers in the tradition of Lord Kelvin's tide predictor of 1876; mechanical, electromechanical, and electrical analyzers (such as the differential analyzer proposed by Charles Babbage in the 1820s, but never completed); electrolytic tanks; resistive papers and elastic membranes to model two dimensional problems; and an incredible array of fire-control computers, extending from the simple computing mechanism of the anti-aircraft acoustic detectors of the 1920s to the complex anti-aircraft director of the 1940s—a machine that consisted of thousands of parts and required several men for its operation. What escapes this already large list are artifacts that defy easy classification because they were extremely unique structures and were uniquely named (such as those used for tidal calculations in the Netherlands or as the Bell Labs isograph) and they may amount into an even more lengthy listing.

The long list of analog computers corresponds to an equally long list of computer architectures. By constitution idiosyncratic, analog computers cannot be defined by reference to shared technical characteristics. This explains why all available technical definitions of the analog computer never went beyond the attempt to define the technical difference between the analog and the digital computer. All things considered, the technical difference between an analog slide rule and a digital desktop machine was through encasing the motion of the latter into a box (blackboxing), which divided the visible computing action into an open or public, and a closed or private (encased) distinction. In general, the analog computer became conceptualized by the visible mechanical and electrical motion that mediated between the input and the output of numerical data whereas by comparison, the digital computer, given that this motion was invisible, became conceptualized by the restricted attention to the input and output of the numerical data themselves.

In 1920, Vannevar Bush of the Massachusetts Institute of Technology (MIT)—a leading figure of interwar computing who later became a founder of postwar science and technology policy—while seeking to connect electric power transmission lines into networks, experimented with machines as complex and fully electrified as an artificial line and with machines as simple and mechanical as a nomogram (or alignment charts). Between 1927 and 1943, Bush supervized the development at MIT of a new spectrum of electromechanical machines to be used in electric power transmission, ranging from his special-purpose network analyzer (an electrical machine that modeled in miniature a large-scale power grid by measuring varying voltages) to the general-purpose differential analyzer (mechanical), completed in 1931 with the aid of his students. The MIT differential analyzer was widely copied in Europe and in America between 1934 and the early 1950s.

How network analyzers became used in other scientific and technical computations, including the computation of subatomic structures by Douglas Hartree's analyzer at Manchester, reinforces the subjectivity of the demarcation between a special and a general-purpose computer. Flow diagrams accumulated by the intensive use of the analog network analyzer were subsumed under the digital computer software (programs) of electric power transmission—an example typical of the continuity between the richness of analog hardware (early twentieth century) and digital software (late twentieth century). Bush's differential analyzer used mechanical integrators, as did Kelvin's tide predictor (though Bush did not know of Kelvin's ideas at the time), torque amplifiers, drive belts, shafts, and gears and was driven by electrical servomotors. Unlike the tide predictors, it could solve differential equations for any application, leading to demand for further machines.

General Electric's Edith Clarke, who started as a "computor" (i.e., a person who produced computations using a desk calculator) before becoming an electrical engineering analyst (i.e., a person who directs computors), developed several unusual electric power transmission slide rules and graphical calculators in the 1920s. The most unusual electric power transmission computers were perhaps those of Vladimir Karapetoff, an electrical engineering professor at Cornell University from

1908–1939. Named after distinguished electric power transmission specialists, the "Heavisidion" and the "Blondelion," Karapetoff's computers may be classified as combinations of kinematic computing mechanisms with linkages, a class of computers that became better known by the work of Antonin Svoboda in the 1940s.

The military counterpart of this civilian line of computer development points to the same diversity. In the military fire control of artillery, the range of what we now call analog computers extended from something as simple as the slide rule to something as complex as an aircraft computing bombsight, or its inverse, a radar-controlled anti-aircraft gun director that directed guns to fire based on target positions of aircraft and correcting for shell speed and flight time. It took Macon Fry a series of seven *Machine Design* articles (September 1945–February 1946) to describe the basic constituents of pre-electronic fire-control computing mechanisms. The Bush differential analyzer at the Moore School of Electrical Engineering at the University of Pennsylvania, built in 1942, was used in World War II ballistics calculations. Another mechanical analog computer, the Norden bombsight, which was used to compute the trajectory of the first atomic bomb in 1945, perhaps qualifies as the most important computer of the twentieth century. Interwar fire-control technology suggests that the role of the state in fostering important changes in computing technology has much deeper roots than canonically assumed. The military contracts awarded to manufacturers of digital calculating machines and punched card machinery to produce analog electromechanical fire-control computers set the stage for the crucial transfer of technology that made IBM's postwar triumph possible. Noticeably, bombsight technology was a national secret, second only to the atomic bomb. From bombsights to analyzers, the pre-1950 ideology of intelligent machines in the history of electromechanical analogs prepared for the acceptance of the electronic computer as a thinking artifact after 1950.

From the General Electric versus Westinghouse competition in renting or selling network analyzers to electric power utilities, to the legendary clash between the Sperry Gyroscope and the Norden bombsights, the business antagonisms of recent decades are prefigured in the history of the development and use of electromechanical analog computers'. Some of the firms involved; for example, Vickers (Europe) and Sperry (U.S.), maintained a presence in their field for many decades. We possess considerable information about a few items, but practically nothing about other equally important artifacts—not to mention that our knowledge barely delves into the knowledge of how to use these artifacts. For example, the development of the aforementioned electric power transmission computers cannot be adequately understood without reconfiguring Clarke's method of "symmetrical components" or Bush's "operational calculus." A better engagement with the history of pre-electronic analogs would doubtless identify many surprising contributors. We have several studies on Bush's network analyzers but very little on the artificial lines of Bush's dissertation advisee, Arthur Kennelly—not to say that we need to learn more about the computers of Kennelly's early employer, Thomas Edison, who computed his first electric networks by relying on miniatures constructed by his Austrian employee, Dr. Herman Claudius.

The analog electronic computer emerged in one of two ways: either by reconfiguring pre-World War II electromechanical analogs, or, as was more common, by the initial configuration of novel electronic computer designs afforded by the post-Word War II cheapening of their constituent electronic components. In the face of rapidly increasing costs for special-purpose software to run the general-purpose electronic digital computer, the costs for which decreased rapidly, crucial advances in fields as important as that of aeronautical design rested almost exclusively throughout the third quarter of the twentieth century, on the development and use of the electronic analog computer.

By the early 1960s, the peak period in the development of the electronic analog computer, its subclasses and uses were so prolific that it took Stanley Fifer a series of four volumes simply to introduce them. Between the late 1940s and the early 1970s, the total production of books and journals on electronic analog computers and associated technologies (e.g., hybrid computers, analog-to-digital and digital-to-analog converters—see Computers, Hybrid) filled rooms of libraries in polytechnic institutions around the world. Now covered by a quarter century of dust, the pages of these publications tell a story that challenges the canonical explanation of the technical inevitability of the victory of the digital over the analog. It is a story that juxtaposes the specific merits of concrete utilization of analog speed to the abstract rhetoric about the customary virtues of the digital accuracy.

All evidence suggests that analog electronic computers became indispensable as controllers of industrial processes (direct analogs). Of the innu-

merable uses of electronic analog computers—in contexts that privileged solutions in real or faster-than-real time—the best studied are those from the military. Rockets, missiles (including intercontinental), and military aircraft developed concurrently with military analog electronic computers. This coevolution became possible by a series of pioneering military projects launched in the late 1940s and the early 1950s. In the U.S., these projects were subcontracted to private firms; in Britain, they were developed at in-house facilities. The result was the eventual development of a rich global network of prospering electronic analog computer manufacturers. The military analog computers were mathematical machines (indirect analogs) used for the solution of the differential equations corresponding to the phenomena to be computed. They were programmed, which means assembled and reassembled, by connecting units or modules of units of alternating or direct current operational amplifiers, used to perform addition, subtraction, integration, and differentiation. The replacement of mechanical amplifiers (the most famous of which was the wheel-and-disk integrator) by electronic versions resulted in the development of electronic differential analyzers (also indirect analogs). The electrical network analyzer, first cousin of the mechanical differential analyzer (a more direct analog), was also restructured by the inclusion of electronic components. The result was a developmental line that concluded in electronic computers similar to Westinghouse's Anacom.

Beginning in the late 1940s, the analog–digital debate continued and escalated into the 1950s. The value of the analog as a tool for demonstration and visualization of computing problems helped it decisively to remain in the market in the 1970s as a tool for educational and instructional purposes. Those who have studied the analog–digital debate from a perspective that is sensitive to the analog side see a social tradeoff between the analog and the digital instead of an inevitable technical evolution from the analog to the digital. Extensions of this debate included, for example, the contrast between the adaptability in use afforded by the hardware malleability of the analog computer versus the generality of use permitted by the software flexibility of digital computer.

The contemporary proponent of the analog seeks to counter the obvious respect to the precision of a solution being paid by a mental worker, mathematician, or accountant, on the one hand, with the pleasure of the "feeling" the solution enjoyed by a manual worker, craftsman or technician, on the other hand. In the first part of the twentieth century, engineers argued that "feeling," "insight," and "intuition" of the problem under computation would be lost in choosing the "invisible," "inner," and "private" workings of the digital desktop machine over the fully "visible," "external," and "public" act of computing with the analog slide rule. In his comparison of analog and digital electronic installations of the 1950s and the 1960s, James Small finds something similar.

Digital computer installations were frequently organized on a closed-shop basis, with computer operators being frequently responsible for scheduling computer usage while users often simply submitted their problem. The fate of its solution was left to numerical analysts and programmers who, first, had no physical insight, and second, need not know how the computer worked. By contrast, analog computer installations usually operated on an open-shop basis, with engineers having direct access to the computer, configuring and reconfiguring it in a process of a computation that was capable of dynamically adjusting to identifications of error sources due to the mathematical model or dysfunctional components. Thus seen, the story of the analog computer is the story of living labor (variable capital) as the other side of automatic machinery (constant capital). As such, the analog versus digital debate of the third quarter of the twentieth century is the new version—that of the electronic era—of a struggle that is as old as the capitalist mode of production.

The historiographical retrieval of the significance of (electronic) hybrid computing, which was a mode of computing that was developed to combine the advantages of the analog and the digital (see Computers, Hybrid), provides further indication of the analog–digital relationship being more historically complex than previously assumed. One may simply start by commenting upon the notoriously problematic identification of analog computer subclasses. For Fifer, the class of analog computers, as distinct, supposedly, from that of digital computers, contained the option of constructing a computer (as opposed to a "simulator-tester", a "trainer" or a "controller") that was indirect-functional (as opposed to direct); compressed-fast time (as opposed to real or extended-slow time); general (as opposed to special); and electronic (as opposed to mechanical, electrical or electromechanical). It is unclear how an electronic general-purpose compressed-time indirect computer differed from an electronic digital computer, which means that in Fifer's classification (1961) the digital electronic computer was actually a subclass of the analog. It is then a

matter of interpretation whether the analog actually lost to the digital or it was simply superseded by the digital.

Historians of the digital computer find that the experience of working with software was much closer to art than science, a process that was resistant to mass production; historians of the analog computer find this to have been typical of working with the analog computer throughout all its aspects. The historiography of the progress of digital computing invites us to turn to the software crisis, which perhaps not accidentally, surfaced when the crisis caused by the analog ended. Noticeably, it was not until the process of computing with a digital electronic computer became sufficiently visual by the addition of a special interface—to substitute for the loss of visualization that was previously provided by the analog computer—that the analog computer finally disappeared.

See also **Calculators, Mechanical and Electromechanical; Computers, Early Digital; Computers, Hybrid**

ARISTOTLE TYMPAS

Further Reading

Aspray, W. Edwin L. Harder and the Anacom: analog computing at Westinghouse. *IEEE Ann. Hist. Comput.* 15, 2, 35–52, 1993.

Bennett, S. *A History of Control Engineering, 1930–1955*. Peter Peregrinus, London, 1993.

Bromley, A. Analog computing devices, in *Computing before Computers*, Aspray, W., Ed. Iowa State University Press, Ames, IA, 1990.

Bush, V. Instrumental analysis. *Trans. Am. Math. Soc.*, 42, 10, 649–669, 1936.

Fifer, S. *Analogue Computation: Theory, Techniques, and Applications, vols. I–IV*. McGraw-Hill, New York, 1961.

Fry, T.C. Industrial mathematics. *Bell Syst. Tech. J.*, 20, 3, 255–292, 1941.

Hartley, M.G. *An Introduction to Electronic Analogue Computers*. Methuen, London, 1962.

Higgins, W.C., Holbrook, B.D. and Emling, J.W. Defense research at Bell Laboratories. *Ann. Hist. Comput.*, 4, 3, 218–244, 1982.

Hopp, P.M. *Slide Rules: Their History, Models, and Makers*. Astragal Press, Mendham, NJ, 1999.

Horsburgh, E. M., Ed. *Modern Instruments and Methods of Calculation: A Handbook of the Napier Tercentenary Exhibition*. Bell, London, 1914.

Jackson, A.J. *Analog Computation*. McGraw-Hill, New York, 1960.

von Jerzerski, D. *Slide Rule: A Journey through the Centuries*. Astragal Press, Mendham, NJ, 2000.

Leffingwell, W.H. *The Office Appliance Manual*. National Association of Office Appliance Manufacturers, 1926.

Lipka, J. *Graphical and Mechanical Computation*. Wiley, New York, 1918.

Karplus, W. *Analog Simulation: Solution of Field Problems*. McGraw-Hill, New York, 1958.

McFarland, S.L. *America's Pursuit of Precision Bombing, 1910–1945*. Smithsonian Institution Press, Washington D.C., 1995.

Mindell, D. *Between Human and Machine: Feedback, Control, and Computing Before Cybernetics*. John Hopkins University Press, Baltimore, 2002.

Owens, L. Vannevar Bush and the differential analyzer: the text and the context of an early computer. *Technol. Cult.*, 27, 1, 1986, 63–95.

Persson, P.-A. Transformation of the analog: the case of the Saab BT 33 Artillery Fire Control Simulator and the introduction of the digital computer as control technology. *IEEE Ann. Hist. Comput.*, 21, 2, 52–63, 1999.

Paynter, H.M. *A Palimpsest on the Electronic Analog Art*. Philbrick Researches, Boston, 1955.

Small, J.S. General-purpose electronic analog computing, 1945–1965. *IEEE Ann. Hist. Comput.* 15, 2, 8–18, 1993.

Small, J.S. Engineering technology and design: the post-Second World War development of electronic analogue computers. *Hist. Technol.*, 11, 33–48, 1994.

Small, J.S. *The Analogue Alternative: The Electronic Analogue Computer in Britain and the USA, 1930–1975*. Routledge, London, 2001.

Tomayko, J.E. Helmut Hoelzer's fully electronic analog computer. *IEEE Ann Hist. Comput.*, 7, 3, 227–240, 1985.

Various. Twentieth century analog machines. *IEEE Ann Hist. Comput.*, Special Issue, 18, 4, 1996.

Wilden, A. *System and Structure: Essays of Communication and Exchange*. Tavistock, London, 1972.

Computers, Early Digital

Digital computers were a marked departure from the electrical and mechanical calculating and computing machines in wide use from the early twentieth century. The innovation was of information being represented using only two states (on or off), which came to be known as "digital." Binary (base 2) arithmetic and logic provided the tools for these machines to perform useful functions. George Boole's binary system of algebra allowed any mathematical equation to be represented by simply true or false logic statements. By using only two states, engineering was also greatly simplified, and universality and accuracy increased. Further developments from the early purpose-built machines, to ones that were programmable accompanied by many key technological developments, resulted in the well-known success and proliferation of the digital computer.

Early digital computers made use of electromechanical relays, adapted from telephone exchanges. Relays were essentially on–off switches, and implementation of binary logic was simply a matter of wiring, with the logical element called the flip-flop being the fundamental component of binary information storage. This was a simple

combination of two basic logic "gates," themselves constructed from an arrangement of switches.

Among the first recorded digital computers was a machine built by a German, Konrad Zuse. In 1936 he constructed a binary calculating machine from electromechanical relays in his living room. This performed binary arithmetic, floating point calculations and program control was by a punched tape, widely used in telegraphs. His work was independent of projects elsewhere in the U.K. and U.S. and did not result in proliferation or market success.

The first round of digital computers in the U.S. also made use of electromechanical relay technology and was seeded by Claude Shannon's 1937 master's thesis at the Massachusetts Institute of Technology (MIT). He described a way of using symbolic logic to improve electrical switching circuits and described an electrical circuit made from relays and switches that could add binary numbers. Since relay technology was widely used in telephone exchanges, it was mature and broadly available. Bell Telephone Laboratories in the U.S. tested ideas on a small scale with their relay interpolator, which consisted of 500 relays and later the ballistic computer, which contained 1300 relays. Their work cumulated in 1944 with an all-purpose computer with 9000 relays and operated by 50 teletype consoles.

International Business Machines (IBM), which had established itself in the office adding and tabulating machine market, took up ideas put forward by Harvard University's Howard Aiken to solve existing problems of "insufficient means of mechanical computation." Under IBM's sponsorship and capitalizing on IBM's existing punched card readers and writers, Aiken's Mark 1 computer (also known as the IBM automatic sequence controlled calculator) was developed at Harvard University with the assistance of a group of IBM engineers and also used electromechanical relay technology. Completed in 1944 at 15.5 meters long and 2.4 meters high, it was considered the first automatic general purpose digital machine. It had four tape readers in which data and program code was fed, and subroutines stored. A more capacious Mark II followed. The machines were used by the U.S. Navy for tackling ballistic problems, one of the recurring needs during wartime for "number crunching."

In the U.K., pioneering efforts in digital computing were made by mathematician Alan Turing, whose idea of a "Universal Turing Machine" helped break the conceptual ground for automated computers tackling mathematical problems. His 1936 paper "On Computable Numbers" theoretically described a machine that could do any calculation that could be done by a human. During World War II, his work with developing computing systems at Bletchley Park was fundamental in cracking German U-boat "Enigma" communications and keeping the Atlantic shipping lanes open, ultimately contributing towards the Allied victory. The purpose-built 1944 "Colossus," built by Thomas Flowers of the Post Office Research Laboratories was built from relays from Post Office telephone exchanges, and operated with a fixed sequence for deciphering the Germans U-boat communications encrypted with Enigma codes.

These machines marked the transition of computers from calculating machines to automated computers. However the bulk, power requirements, and the delay time in the operation of relays were performance-limiting factors.

The development and application of the electronic vacuum tube (also known as the valve) as a switching unit in digital computers permitted vastly improved performance compared to relay technology. The first unit to employ large-scale use of valves for digital computing was the ENIAC (electronic numerical integrator and computer) which consisted of over 18,000 valves (see Figure 13). Built by John Mauchly and John Prosper Eckert at the Moore School of Electrical Engineering at the University of Pennsylvania to compute ballistic tables for the U.S. Army, it was only completed after the close of the war in late 1945. Although it could only store 20 binary words, its unprecedented calculation speed attracted attention from scientists and professionals, in particular a key member of the Los Alamos Manhattan Project for the atomic bomb, mathematician John von Neumann. The ENIAC was used for other scientific purposes such as weather predicting and other fluid dynamic calculations, including calculations for the fusion hydrogen bomb. However, programming the ENIAC meant setting arrays of function switches and a maze of patch cables. This was time consuming and meant that it could not be used for computing whilst it was being programmed. This drove the development of the stored program computer, whereby the computational instructions could be prepared on tape and loaded quickly into memory.

The EDVAC (electronic discrete variable calculator) was the first stored program electronic digital computer and was also built by the Moore School under contract to the U.S. Army. It was highly influenced by von Neumann's work in

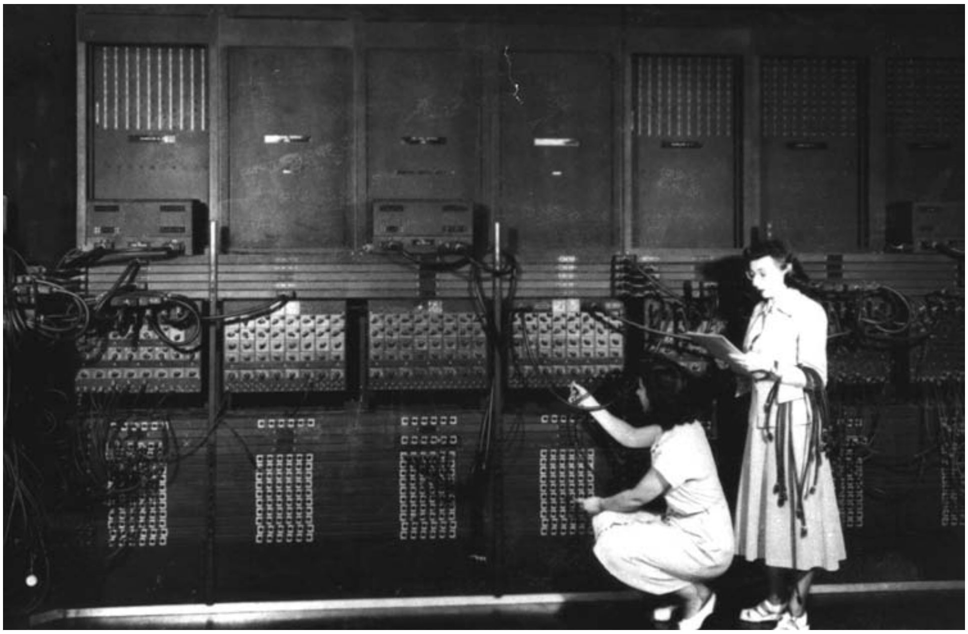


Figure 13. Two women wiring the right side of the ENIAC with a new program. [*U.S. Army Photo from the archives of the ARL Technical Library.*]

formal logics since he provided critical early direction into the logical structure and design of the computer in 1945, together with Mauchly and Eckert. Von Neumann abstracted the logical design from the hardware, and introduced the concept of the "stored program" whereby programming instructions are stored as numbers in memory alongside numerical data. The EDVAC, installed at the Ballistic Research Laboratory at Aberdeen Proving Ground, MD in 1949 but not fully complete until 1952, also introduced mercury delay lines for primary memory. Developed by John Eckert for the project, these devices increased storage density by a factor of 100 over valve devices.

The ideas and work of the Moore School rapidly proliferated thereafter. A summer course on computer design based on the von Neumann concept was run by that institution in 1946, and was attended by British and American professionals. Also in 1946, von Neumann, Herman Goldstine and Arthur Burks published a comprehensive report of their work at Princeton's Institute of Advanced Study Electronic Computer Project, where they had established themselves following their involvement with the Moore School. The report detailed the operation and architecture of their work on digital computers, and has been described as the blueprint for "modern" digital computing. The authors' insistence that it be placed in the public domain, as is the tradition with scientific publishing, was certainly an important factor in the spread and adoption of this knowledge. The design proposed for the EDVAC became the basis for several machines, including in England, the electronic delay storage automatic computer (EDSAC) built by Maurice Wilkes at the Mathematical Laboratory at Cambridge University completed in 1949, and the Manchester Mark I.

The first digital computing machine to store both program and data in random-access memory (RAM) was Manchester University's Mark I, built by Geoff Toothill, Frederick Williams and Tom Kilburn in 1949, using cathode-ray tubes for fully electronic memory. The Mark I is considered by many to be the first stored program computer, since it was in use before the EDVAC was fully operational.

The Whirlwind was developed at the Servomechanism Laboratory at MIT in Boston, and was sponsored by the U.S. military for real-time simulation of the effect of engineering changes on aircraft behavior. The project began in 1946 and while it drew from the work at Princeton and Moore Schools, the designers' background in automatic control and electrical engineering gave it a certain technological uniqueness. The Whirlwind project resulted in some key technological developments such as magnetic core memory and the parallel

synchronous method of handling information. Developed by Jay Forrester in 1949, magnetic core memory was an important step in miniaturization and speed of computer memory. The Whirlwind was only commissioned in 1958, finding use as the control and prediction computer for the SAGE air defense system. This relied on real-time communications to remote radar stations for which a key digital communication technology was developed—the modem. However, it has been argued that Whirlwind's time was over before it ever began: the system had become redundant by the Cold War proliferation of intercontinental ballistic missiles.

The UNIVAC is regarded as the first commercial digital computer. It was designed and built by J. Presper Eckert and John Mauchly who had worked extensively on the engineering of ENIAC project. The acronym stems from "universal automatic computer," the intention being that it would have universal appeal to scientists, engineers and business. Although the core functional "organs" were similar to the EDVAC, it was a more robust machine and required less maintenance. The UNIVAC incorporated some key improvements that vastly speeded up data processing which included a magnetic tape system for secondary memory storage and a data buffering mechanism to the delay line primary memory storage. The first commercial installation of a UNIVAC was at the U.S. Census Bureau in 1951. LEO (Lyons electronic office) however went into commercial office action a few months before UNIVAC. LEO, based on Cambridge University's EDSAC, was used by Lyons & Company Limited Bakery for data processing. With interest from other companies, Lyons later sold LEO II computers to other British firms.

IBM responded to UNIVAC in 1952 with the Model 701, which was technologically similar to the UNIVAC with the exception that it featured cathode-ray tube memory. This was a technology appropriated from the television industry, the phosphorescence of the screen providing the ability to store binary signals. Another key feature was the use of a magnetic oxide-coated drum for secondary storage (the predecessor of the hard disk) as well as a lightweight plastic tape tertiary memory. This had a lower inertia than the metal tapes of the UNIVAC and sped up tape operations considerably. Due to IBM's massive market presence and infrastructure, the 701 was a considerable market success.

Following the success and proliferation of these computers in business and scientific research organizations from the early 1950s, a plethora of startups presented the global market with a diverse range of digital computers. These companies were initiated by key staff members from the pioneering organizations, such as Seymour Cray and William Norris. Besides technological improvements in processing speed, primary and mass storage memory density, architecturally digital computers did not diverge from the classic "von Neumann architecture." Furthermore, military-sponsored research in the U.S. produced key advances such as the transistor, the integrated circuit and computer networking (from which the Internet grew) although personal computing was more of a civilian initiative. Larger demand for computers in all spheres of society spawned a global industry, which ultimately drove down prices and increased miniaturization, and resulted in the digital computer becoming ubiquitous in most industrialized societies by the turn of the century.

See also **Calculators, Mechanical and Electromechanical; Computer Memory, early; Computers, Analog; Computers, Mainframe; Encryption and Code Breaking; Vacuum Tubes/Valves**

BRUCE GILLESPIE

Further Reading

Aspray, W. *John von Neumann and the Origins of Modern Computing*. MIT Press, Cambridge, MA, 1990.

Ceruzzi, P.E. *A History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Eames, C. and Eames, R. *A Computer Perspective, Background to the Computer Age*. Harvard University Press, Cambridge, MA, 1973.

Goldstine, H.H. *The Computer: From Pascal to von Neumann*. Princeton University Press, Princeton, 1972.

Hodges, A. *Alan Turing: The Enigma*. Burnett Books, London, 1983.

Hughes, T.P. *Rescuing Prometheus: Four Monumental Projects that Changed the Modern World*. Vintage Books, 1998.

Irvine, M.M. Early digital computers at Bell Telephone Laboratories. *Ann. Hist. Comput*., 23, 3, 22–42, 2001.

Computers, Hybrid

Following the emergence of the analog–digital demarcation in the late 1940s—and the ensuing battle between a speedy analog versus the accurate digital—the term "hybrid computer" surfaced in the early 1960s. The assumptions held by the adherents of the digital computer—regarding the dynamic mechanization of computational labor to accompany the equally dynamic increase in computational work—was becoming a universal ideology. From this perspective, the digital computer

justly appeared to be technically superior. In introducing the digital computer to social realities, however, extensive interaction with the experienced analog computer adherents proved indispensable, especially given that the digital proponents' expectation of progress by employing the available and inexpensive hardware was stymied by the lack of inexpensive software. From this perspective—as historiographically unwanted it may be by those who agree with the essentialist conception of the analog–digital demarcation—the history of the hybrid computer suggests that the computer as we now know it was brought about by linking the analog and the digital, not by separating them. Placing the ideal analog and the ideal digital at the two poles, all computing techniques that combined some features of both fell beneath "hybrid computation"; the designators "balanced" or "true" were preserved for those built with appreciable amounts of both.

True hybrids fell into the middle spectrum that included: pure analog computers, analog computers using digital-type numerical analysis techniques, analog computers programmed with the aid of digital computers, analog computers using digital control and logic, analog computers using digital subunits, analog computers using digital computers as peripheral equipment, balanced hybrid computer systems, digital computers using analog subroutines, digital computers with analog arithmetic elements, digital computers designed to permit analog-type programming, digital computers with analog-oriented compilers and interpreters, and pure digital computers.

There were many types of analog-to-digital converters, including raster, chronometric, rotating drum, null detector, direct counting, incremental steps, star wheel, fixed interval, and commutator and brush, which transformed continuous analog information into discrete numerical data. Their common feature was an accurate standard of spatial or temporal measurement to set the analog information into digital form.

Given the tendency towards digitalization, the literature on analog-to-digital converters was more expansive than its opposite: digital-to-analog converters. In addition to converters, bilateral operation could rely on multiplexers and demultiplexers to permit the converters to be time shared, and to hold devices, buffers, timing, and control circuitry. The increasing demand for hybrid interface equipment brought about display devices (plotters, scopes, etc.) that accepted digital outputs and produced analog-appearing inputs, and conversely, similar artifacts (sketchpads, tablets, and light pens) that accepted analog-type inputs and converted them to digital signals suitable for computer input.

The way analog and digital computers could be used together was classified into two broad categories: unilateral operation in which information flew across the interface between the analog and the digital sections in one direction, and bilateral operation in which the flow across this interface was in both directions. Noticeably, in bilateral operation the analog or the digital computer could be regarded as playing the part of a complex and elegant input or output device. It therefore follows that claiming that the digital won the analog–digital debate is similar to arguing that the history of hybridization ended with the victory of a unilateral hybrid computer in which the information flowed only to the digital section.

Early hybridization, called analog–digital conversion techniques, suffered from the lack of adequate software, inadequacies of vacuum-tube linkage elements, and the unsuitability of available digital computers for online operation. Initially, most hybrid computer systems employed large digital computers, such as the IBM 7094 or the Univac 1103. During the 1960s, some smaller digital such as the Scientific Data Systems 930 were also used. Eventually, hybrid computers were used in sampled-data system simulation, random process simulation, optimization, simulation of distributed parameter systems, studies in guidance and control of missiles and space vehicles, simulation of man-machine systems, and numerous other application contexts. Hybrid speed was, for example, advantageous in problems with random variables where a large number of solutions were required for statistical analysis.

One notable development was the configuration of a hybrid computing Monte Carlo technique, which permitted accumulation of statistics taken from thousands of fast analog-computer runs. The hybridization of electronic analog and digital computers was also undertaken in order to overcome computational problems in the design and development of intercontinental ballistic missiles (ICBMs). Combining the analog's speed and the digital's accuracy through conversion was first tried at Convair Astronautics in 1954 and at the Ramo-Woolridge Corporation in 1955. Both combinations were undertaken to create the Atlas ICBM system. In the late 1950s and early 1960s, several experimental hybrids undertaken by Electronic Associates, Inc. (EAI) and Packard-Bell were built by linking analog and digital through interfaces. In the mid-1960s, Comcor/

Astrodata and EAI introduced fully transistorized analog computers and highly integrated hybrid computer systems began to be manufactured commercially.

By the 1960s, hybrid equipment was used by nearly every firm and government research and test facility working on missile design and other aerospace applications. Space exploration provided the impetus for the further development of the hybrid computer. James Small argues that the manned space program took over where the problems of cost, complexity, and hazards in ICBM development and simulation left off. Including the astronaut in the control loop increased computing complexity. Large, fully transistorized true-hybrids were used in the Gemini and the Apollo programs and in the Saturn-V rocket system. Hybrids were employed in a wide range of space applications, including the simulation of space vehicle docking, engine and navigation system design, and astronaut instruction.

By the late 1960s, more than ten firms were manufacturing hybrid computer systems, and while sales of analog computers that were not fully hybrid-compatible were decreasing, sales of hybrid computers were increasing. Sales of hybrids peaked at approximately $50 million per annum in the late 1960s, falling substantially when aerospace budgets were eventually reduced in the 1970s. They remained, however, between $25–$30 million through 1975. As late as 1989, there were 300 large installations of hybrids worldwide. The decline of the hybrid cannot, however, be explained by economic necessity alone. Small notes that by 1970 at least 75 percent of all simulations were performed on digital computers, even after studies showed that the hybrid computer offered speed advantages greater than the fastest digital computers by as much as 100 to 1. In army material control studies, advanced hybrid computer systems (fourth generation hybrids) showed potential savings of 30 to 1 by using hybrids, amounting to a potential market of $300 million in the U.S. alone. All specialists agreed that hybridization had resulted in increased computing hardware utilization. For some, hybridization was also a healthy response to an unfortunate specialization of interest and skills by computer engineers, which was the outcome of the fanatical separation between analog and digital computers. Hybridization helped the efficient transformation of analog schematics into digital flow charts, and the creation of utility libraries meant less duplication of programming work. Others, however, feared that hybridization would bring about a dependence on automatic programming techniques that could isolate engineers from the computer, thereby replacing the open-shop computing environments of the analog computer by the closed-shop computing environments of the digital computer.

See also **Computers, Analog; Computers, Early Digital**

ARISTOTLE TYMPAS

Further Reading

Bekey, G.A. and Karplus, W.J. *Hybrid Computation.* Wiley, New York, 1968.

Fifer, S. *Analogue Computation: Theory, Techniques, and Applications*, vol. II. McGraw-Hill, New York, 1961.

Hydman, D. E. *Analog and Hybrid Computing*. Pergamon Press, Oxford, 1970.

Korn, G.A. and T. M. Korn. *Electronic Analogue and Hybrid Computers*, 2nd edn. McGraw-Hill, New York, 1972.

McLeod, J. *Simulation: The Dynamic Modeling of Ideas and Systems with Computers.* McGraw-Hill, New York, 1968.

Small, J.S. General-purpose electronic analog computing, 1945–1965. *IEEE Ann. Hist. Comput.*, 15, 2, 8–18, 1993.

Small, J.S. Engineering technology and design: the post-second world war development of electronic analogue computers. *Hist. Technol.*, 11 33–48, 1994.

Small, J.S. *The Analogue Alternative: The Electronic Analogue Computer in Britain and the USA, 1930–1975.* Routledge, London, 2001.

Susskind, A.K., Ed. *Notes on Analog–Digital Conversion Techniques.* Wiley, New York, 1957.

Computers, Mainframe

The term "computer" currently refers to a general-purpose, digital, electronic, stored-program calculating machine. The term "mainframe" refers to a large, expensive, multiuser computer, able to handle a wide range of applications. The term was derived from the main frame or cabinet in which the central processing unit (CPU) and main memory of a computer were kept separate from those cabinets that held peripheral devices used for input and output.

Computers are generally classified as supercomputers, mainframes, minicomputers, or microcomputers. This classification is based on factors such as processing capability, cost, and applications, with supercomputers the fastest and most expensive. All computers were called mainframes until the 1960s, including the first supercomputer, the naval ordnance research calculator (NORC), offered by International Business Machines (IBM) in 1954. In 1960, Digital Equipment Corporation (DEC) shipped the PDP-1, a computer that was much smaller and cheaper than a mainframe.

These more affordable machines, called minicomputers, became increasingly popular in the commercial sector after the introduction of the PDP-8 in 1964 and the PDP-11 in 1970, and widely increased the base of consumers and applications. Microcomputers were introduced in the 1970s, including such machines as the Kenback I, the Altair, and the Apple; these further broadened the computer market (see Computers, Personal; Computers, Supercomputers).

All computers contain processors, memories, and peripherals. The CPU manipulates instructions and data that are stored in its registers and in main memory. The CPU and memory consist of electronic components, allowing current processing speeds in the billions of instructions per second. Nanotechnology is being researched for even faster speeds. All storage within the CPU and main memory is currently volatile; if power is removed, the stored values are lost. The peripheral devices typically have mechanical components that limit their speed. These include nonvolatile secondary storage devices, which hold programs and data for input and output on demand, as well as devices used to interface with humans, such as monitors, printers, and terminals. Other devices provide for computer-to-computer interactions, such as those used by networks. Special-purpose processors perform various specialized functions, such as floating-point operations or input and output of data.

Early electronic computing machines were analog, with computed values analogous to input values. Analog machines, first designed by Vannevar Bush at MIT in 1927–1935, were widely used during World War II for calculating ballistics tables for weapons systems. They were soon overshadowed by digital computers, in which all data and control information (programs) are encoded with discrete values. A great many inventors contributed to the development of the various components of modern computers. Following are some highlights of this process.

In 1940, George Stibitz demonstrated a communications cable between his complex calculator and a teleprinter, which was a forerunner of modern digital communication ports. Konrad Zuse, in association with Helmut Schreyer, is credited with building the first working general-purpose program-controlled computing machine, the Z3, in 1941. John Atanasoff and Clifford Berry developed an electronic digital computing machine between 1939 and 1942; their ABC machine utilized vacuum tubes (valves) for switches. Influenced by the work of George Babbage, between 1939 and 1943 Howard Aiken led the development of the Harvard Mark I, a program-controlled digital calculator with electromechanical components. By 1943, under the direction of Tommy Flowers, the Colossus, a large-scale, electronic, digital computing machine was developed for use in deciphering German code during World War II (see Computers, Early Digital). John Mauchly and J. Presper Eckert designed a general-purpose digital electronic machine, the ENIAC (electronic numerical integrator and computer), which was completed in 1946. Between 1946 and 1952, John von Neumann spearheaded the development at Princeton of the Institute of Advanced Study (IAS) machine, which stored its programs in memory together with data. The late 1940s saw the introduction of Maurice Wilkes's EDSAC (electronic delay storage automatic computer) at Cambridge University and the EDVAC (electronic discrete variable calculator) at the Moore School, University of Pennsylvania, both with stored programs in a memory of cathode-ray tubes and mercury delay lines. In 1948, IBM marketed the 604 Electronic Calculating Punch Card Machine, with input and output data stored on punched cards and programs stored on a plugboard. Also in 1948, Geoff Toothill, Frederick Williams and Tom Kilburn built a forerunner, called "Baby," for Manchester University's Mark I, the first operational stored-program digital computing machine. In Sydney, Australia, Maston Beard and Trevor Pearcey supervised the building of the Council for Scientific and Industrial Research Computer (CSIRAC) between 1949 and 1951. In 1951, computers became commercially available, with Ferranti's Mark I (based on the Manchester Mark I) and Eckert's and Mauchly's Computer Corporation's (EMCC) UNIVAC I. In the 1950s, the IBM 702 and 704 were developed for business as well as scientific applications. Other early computer manufacturers included Burroughs (Atlas or AN/GSQ-33 Computer, later merged with Sperry-Rand to form Unisys), Remington Rand (acquired EMCC, later merged into the Sperry-Rand Corp.), Texas Instruments (first commercial use of silicon transistors in computers), Fujitsu (Japan's first commercial computer), the Dutch PITT (ZERO and ZEBRA), and Olivetti (Italy's first commercial computer).

IBM became the leader of the mainframe computer market, first with its series 7000 line, and then, overwhelmingly, when it introduced the System/360 family of computers in 1964. The System/360 was a line of computers of increasingly greater size, speed, and number of input and output ports, with a corresponding increase in

cost. All the System/360 computers had the same instruction set, so that software systems developed for machines at the low end of the market could also execute on system upgrades. They featured integrated circuitry, a main memory that was unusually large for the time, and microprogramming. Other innovations included the introduction of different instruction sets for different data types, support for multiprogramming, and the pipelining of instructions. These features were widely influential on the manufacture of successive mainframes, and IBM has continued to dominate the mainframe market to this day (see Computers, Uses and Consequences). Other current mainframe manufacturers include Fujitsu-Siemens, Hitachi, and Unisys.

Mainframes once each filled a large room, cost millions of dollars, and needed a full maintenance staff, partly in order to repair the damage caused by the heat generated by their vacuum tubes. These machines were characterized by proprietary operating systems and connections through dumb terminals that had no local processing capabilities. As personal computers developed and began to approach mainframes in speed and processing power, however, mainframes have evolved to support a client/server relationship, and to interconnect with open standard-based systems. They have become particularly useful for systems that require reliability, security, and centralized control. Their ability to process large amounts of data quickly make them particularly valuable for storage area networks (SANs). Mainframes today contain multiple CPUs, providing additional speed through multiprocessing operations. They support many hundreds of simultaneously executing programs, as well as numerous input and output processors for multiplexing devices, such as video display terminals and disk drives. Many legacy systems, large applications that have been developed, tested, and used over time, are still running on mainframes.

See also **Computer Memory, Early; Computer Networks; Computers, Uses and Consequences; Processors for Computers; Systems Programs**

TRUDY LEVINE

Further Reading

Ceruzzi, P. *A History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Cortada, J.W. *Historical Dictionary of Data Processing Technology*. Greenwood Press, New York, 1988.

Gschwind, H.W. and McCluskey, E.J. *Design of Digital Computers*. Springer Verlag, New York, 1975.

Push, E.W. *Building IBM: Shaping an Industry and its Technology*. MIT Press, Cambridge, MA, 1995.

Stallings, S. *Computer Organization and Architecture*, 5th edn. Prentice Hall, 2000.

Tanenbaum, A. *Structured Computer Organization*, 4th edn. Prentice Hall, 1999.

Computers, Personal

A personal computer, or PC, is designed for personal use. Its central processing unit (CPU) runs single-user systems and application software, processes input from the user, sending output to a variety of peripheral devices. Programs and data are stored in memory and attached storage devices. Personal computers are generally single-user desktop machines, but the term has been applied to any computer that "stands alone" for a single user, including portable computers.

The technology that enabled the construction of personal computers was the microprocessor, a programmable integrated circuit (or "chip") that acts as the CPU. Intel introduced the first microprocessor in 1971, the 4-bit 4004, which it called a "microprogrammable computer on a chip." The 4004 was originally developed as a general-purpose chip for a programmable calculator, but Intel introduced it as part of Intel's Microcomputer System 4-bit, or MCS-4, which also included read-only memory (ROM) and random-access memory (RAM) memory chips and a shift register chip. In August 1972, Intel followed with the 8-bit 8008, then the more powerful 8080 in June 1974. Following Intel's lead, computers based on the 8080 were usually called microcomputers.

The success of the minicomputer during the 1960s prepared computer engineers and users for "single person, single CPU" computers. Digital Equipment Corporation's (DEC) widely used PDP-10, for example, was smaller, cheaper, and more accessible than large mainframe computers. Timeshared computers operating under operating systems such as TOPS-10 on the PDP-10—co-developed by the Massachusetts Institute of Technology (MIT) and DEC in 1972—created the illusion of individual control of computing power by providing rapid access to personal programs and files. By the early 1970s, the accessibility of minicomputers, advances in microelectronics, and component miniaturization created expectations of affordable personal computers.

Innovation in software design during the 1960s and 1970s also shifted attention from computers as large calculating machines to their potential use as

technologies of personal productivity. Work at Douglas Engelbart's Augmentation Research Center at the Stanford Research Institute (SRI), beginning in 1961; David Evans's and Ivan Sutherland's Computer Science Laboratory at the University of Utah and their own company, Evans & Sutherland, founded in 1968; and Xerox's Palo Alto Research Center (PARC), founded in 1969, resulted in prototypes of graphical user interfaces such as Engelbart's NLS software and PARC's Alto, a desktop workstation. The Xerox Star (1981) showed that personal workstations equipped with powerful graphical user interfaces, networking, and office productivity software could be manufactured, but due primarily to its high cost, it was a commercial failure.

The first microcomputers of the mid-1970s, many little more than component kits, were inexpensive and featured simple designs that invited experimentation. The Altair 8800, manufactured by Micro Instrumentation Telemetry Systems (MITS), was the first, making its debut as the cover story of the January 1975 issue of *Popular Electronics*. Encouraged by the article to mail-order kits and build their own computers, hobbyists gathered to share information about microcomputers in clubs such as the Homebrew Computer Club in Palo Alto, California, founded in March 1975. The "computer liberation" philosophy of writers such as Ted Nelson, author of *Computer Lib/Dream Machines* (1974), inspired them to learn about microcomputers as a way of steering computing power "to the people."

Despite the enthusiasm it unleashed, the Altair was not a powerful computer. Ed Roberts, the founder of MITS, was poorly prepared for its success. He realized the need for supporting the claims in *Popular Electronics*, especially the need for software. He hired Bill Gates and Paul Allen—then college students—to write a version of the BASIC (begineer's all-purpose symbolic instruction code) programming system for the Altair; they completed the work in February 1975 after establishing a company, Microsoft, to license the software to him. Dozens of new companies emerged between 1975 and 1977 to manufacture microcomputers with many differences in hardware and software. Many of them were built on the 100-pin bus known as S-100, which had been used by the Altair and the IMSAI 8080 (1976) to provide a channel for the microprocessor to communicate with hardware devices, including third-party add-ons such as Cromemco's Dazzler graphics board (1976). However, it was not a reliable standard. In 1977, Gary Kildall completed a version of the CP/M operating system with a basic input/output system (BIOS) that could be written for each new machine, while the rest of the operating system did not need to be changed. CP/M became an informal software standard for computers that did not rely on proprietary graphics displays, and "CP/M computer" became the generic term for these microcomputers.

Apple Computer was founded in 1976 by Steve Jobs and Stephen Wozniak in order to sell Wozniak's elegantly designed Apple I microcomputer. They launched the Apple II computer at the first West Coast Computer Fair in 1977. Like CP/M computers, the Apple II integrated features such as disk storage with an expandable bus architecture that created an after-market for cards fitting into expansion slots. Sales of the Apple II benefited from new software available for it, such as Daniel Bricklin's and Robert Frankston's VisiCalc, a spreadsheet program in great demand. Jobs worked with engineers, retailers and customers to define personal and business markets for the Apple II, establishing Apple Computer as a leader of the new microcomputer industry. Competitors, such as Radio Shack and Commodore Business Machines followed similar strategies in seeking a customer base beyond the Homebrew hobbyists.

The microcomputer industry rapidly matured in the early 1980s. In August 1981, IBM launched the IBM PC after more than a year of development under the codename Acorn. Its designers utilized Intel's new 8088 microprocessor, beginning the transition to a new generation of more powerful, 16-bit microcomputers. Despite this, the PC was technologically conservative. The MS-DOS operating system written by Microsoft for the IBM PC differed little from CP/M, and the machine itself offered a familiar array of hardware. Yet, the IBM PC, merely by putting IBM's stamp of approval on the "personal computer," significantly expanded the business market for microcomputers.

Graphical user interfaces (GUI) began to appear on personal computers in the early 1980s. At Apple, Jef Raskin proposed a networked personal computer in 1979 utilizing the style of GUI introduced at PARC, just as the company was formulating strategy for a new generation of computers to succeed the Apple II/III line. Raskin's proposal influenced development of the Lisa (1983), a failure in the marketplace. By then, Jobs had taken control of Raskin's Macintosh project, which produced a more economical and compactly designed version of this technology. The Macintosh was introduced in January 1984 with a splashy marketing campaign that depicted it as a

personal computer "for the rest of us," contrasting its innovative design to IBM's staid business-oriented computer.

Despite the fact that GUI interfaces such as VisiCorps VisiOn (1983) and Microsoft Windows (1985) were available for the IBM PC, IBM and Apple by the mid-1980s established alternatives to personal computing. On the one hand, The Macintosh was a closed, tightly controlled architecture emphasizing usability. On the other, the open architecture of the PC using MS-DOS gave manufacturers flexibility in hardware design, with the BIOS linking their machines to the de facto operating system "standard." The evolution of these PC "clones" tracked progress in Intel microprocessors, beginning with the fully 16-bit 8086 chip and continuing through the Pentium processor line, or followed versions of Microsoft's MS-DOS and Windows. The companies that produced them (Compaq, Dell, etc.) paced the phenomenal growth of the PC industry during the 1980s and 1990s.

See also **Computers, Mainframe; Computers, Uses and Consequences; Electronic Communications; Software Application Programs**

HENRY LOWOOD

Further Reading

Bardini, T. *Bootstrapping: Douglas Engelbart, Coevolution, and the Origins of Personal Computing*. Stanford University Press, Stanford, 2000.

Ceruzzi, P.E. *A History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Friedewald, M. *Der Computer als Werkzeug und Medium: Die geistigen und technischen Wurzeln des Persona Computers*. Verlag für Geschichte der Naturwissenschaften und der Technik, Berlin, 1999.

Linzmayer, O.W. *Apple Confidential: The Real Story of Apple Computer, Inc*. No Starch Press, San Francisco, 1999.

Freiberger, P. and Swaine, M. *Fire in the Valley: The Making of the Personal Computer*. Collector's Edition. McGraw-Hill, New York, 2000.

Goldberg, A., Ed. *History of Personal Workstations*. ACM Press, New York, 1988.

Levy, S. *Hackers: Heroes of the Computer Revolution*. Dell, New York, 1984.

Levy, S. *Insanely Great: The Life and Times of Macintosh, the Computer that Changed Everything*. Penguin, New York, 1994.

Computers, Supercomputers

Supercomputers are high-performance computing devices that are generally used for numerical calculation, for the study of physical systems either through numerical simulation or the processing of scientific data. Initially, they were large, expensive, mainframe computers, which were usually owned by government research labs. By the end of the twentieth century, they were more often networks of inexpensive small computers. The common element of all of these machines was their ability to perform high-speed floating-point arithmetic—binary arithmetic that approximates decimal numbers with a fixed number of bits—the basis of numerical computation.

Almost all the early electronic computers were called supercomputers or supercalculators. Two early machines, the Massachusetts Institute of Technology (MIT) Whirlwind (1951) and the International Business Machine (IBM) Stretch (1955) achieved their performance using innovative circuitry. The Whirlwind pioneered the use of magnetic core memory, while the Stretch used high-speed transistors. Though subsequent supercomputers used even more advanced electronics technology, most achieved their speed through parallelism, by performing multiple calculations simultaneously.

Numerical weather prediction, one of the standard problems for supercomputers, illustrates the parallel nature of many scientific calculations. The basic analysis of the earth's weather system, done in the 1920s, imposed a grid upon the surface the earth and developed a series of differential equations to describe how the weather changed at each point of the grid. The equations described such things as the absorption of solar energy, the dissipation of the wind, the movement of humidity, and the fluctuations in temperature. Though some values are passed from point to point, many of the calculations can be done in parallel. This can be best appreciated by comparing two points on opposite sides of the globe. As the weather in London has little connection to the weather in Tahiti, the calculations for the two locations can be done at the same time. Other kinds of problems that shared this kind of geometric parallelism include stress analyses of structures, the processing of graphical images, the simulation of fluid or airflow, and the simulation of nuclear explosions. This last problem was crucial to the evolution of supercomputers. Many supercomputers were developed for the U.S. Defense Department in the 1960s and 1970s. The U. S. nuclear research laboratories helped to test these machines, debug them, and create software for them.

Early attempts to exploit parallelism produced supercomputers called vector machines. Vector processors can perform the same instruction repeatedly on a large amount of data. These

machines are often associated with the designer Seymour Cray. First employed by Sperry-Rand, he helped to found two supercomputer manufacturers: Control Data Corporation (CDC) and Cray Research. His first parallel machine was the CDC 6400, which was announced in 1964. It was a "look-ahead" machine—the elements of the machine could work autonomously and its control unit could execute instructions out of sequence. If it encountered an instruction that needed circuits that were currently in use, it would jump to the next instruction and try to execute it. He expanded this idea in his next two machines, the CDC 7600 and the Cray I.

The vector processor of the Cray I exploited parallelism within floating-point arithmetic. As floating-point numbers are recorded in scientific notation, the fundamental operations, such as addition, subtraction, multiplication, and division are more complicated than the arithmetic of integers. Floating-point addition has four separate steps and requires three separate additions or subtractions. Cray devised logic circuits to perform each step of a floating-point operation. He then arranged these circuits in logical assembly lines, similar in concept to the assembly lines of an automobile factory. The assembly line for addition had four stages, each corresponding to one of the steps. As data passed down these lines, they stopped at each stage to have part of the floating-point operation performed. Once the pipe was filled with data and was processing it, it would produce additions four times faster than a processor that did one complete floating-point addition at a time.

Vector processors were capable of great speeds. If all parts of the Cray I were operating, it would produce 133 million floating-point operations per second, or mega-FLOPS. Yet, the problems of achieving this speed illustrated the software problems for supercomputer designers. Unless a program made heavy use of the vector processor, the machine ran at a fraction of its top speed, an observation known as "Amdahl's Law." The vector hardware could be a thousand times faster than the main processor or even a million times faster, but if the program used it only 5 percent of the time, the supercomputer would run only 5 percent faster than a conventional machine because the supercomputer would always be waiting for the main processor to complete its work. As a consequence, programmers radically learned to rewrite their code in order to keep the vector processor fully occupied, a process known as "vectorization." They commonly discovered that vectorized programs altered the order of basic operations and required substantially more memory to hold intermediate results. Vectorization could be a time-consuming process and difficulty to justify unless the program was run repeatedly. However, by the early 1980s, software developers had created optimizing FORTRAN (from "formula translation," a computer language) compilers that automatically rewrote programs for a vector machine.

During the late 1980s, supercomputer research began to shift from large vector machines to networks of smaller, less expensive machines. These machines promised to be more cost effective, though computer scientists often discovered that these clusters or constellations or processors could be more difficult to program than vector machines. They experimented with different configurations of processors, the extent to which they shared common memory, and the way in which they were controlled. Some machines, such as the Burroughs Scientific Processor, used a single stream of instructions to control an array of processors that worked from a single memory. At the other extreme, the Cray X-MP had independent processors, each of which had large dedicated memories. Between the two, fell machines such as the Connection Machine. It consisted of a large number of processors, each of which had a small amount of independent memory. An important step in this work came in 1994 when a group of NASA Beowulf project. Beowulf machines could be assembled from commercial processors using standard network hardware. Few programmers could utilize the full power of these machines, but these were so inexpensive that many were able to use them. The original NASA Beowulf cost $40,000 and had the potential of computing one billion floating-point operations per second.

With the advent of inexpensive supercomputers, these machines moved beyond the large government labs and into smaller research and engineering facilities. Some were used for the study of social science. A few were employed by business concerns, such as stock brokerages or graphic designers.

See also **Computer Modeling; Computers, Mainframe; Processors for Computers**

David Grier

Further Reading

August, M.C., Brost, G.M., Hsiung, C.C. and Schiffleger, A.J. Cray X-MP: the birth of a supercomputer. *Computer*, 22, 1 45–52, 1989.

Bell, G. and Gray, J. What's next in high-performance computing? *ACM Comm.*, 45, 2, 91–95, 2002.
Cybenko, G. and Kuck, D. Supercomputers-revolution or evolution? *IEEE Spect.*, 29, 9, 39–41, 1992.
Frenkel, K.A. Special issue on parallelism. *ACM Comm.*, 29, 12, 1986.
Hord, M. *Understanding Parallel Supercomputing*. Institute of Electrical and Electronics Engineers, New York, 1998.
Murray, C. *The Supermen: The Story of Seymour Cray and the Technical Wizards Behind the Supercomputer*. Wiley, New York, 1997.
Padmanabhan, K. Cube structures for multiprocessors. *ACM Comm.*, 33,1, 1990.

Computers, Uses and Consequences

Towards the close of the last century the computer was claimed to be the most revolutionary artifact of twentieth century technology. It transformed business and industrial production, engineering and the sciences, as well as everyday life, and brought about a "computer revolution," a "computer age," an "information age," and an "Internet age," as many observers noted. It would therefore appear that the economy and the society of the industrial nations are more and more based on computer technology. Added to this is the ongoing debate over the risks and negative consequences of the increasing ubiquity of computer technology in our society; for example, the effects of computer technology on employment and data privacy protection, and the case of the risks of using software for the control of technological systems such as nuclear weapons systems. The overall picture shows that on the one hand, the computer seems to be a machine that has repeatedly stirred up illusions and visions in society since its invention (e.g., the vision of computers as intelligent machines), and on the other hand, people often fear the consequences of its uses.

The modern computer—the (electronic) digital computer in which the stored program concept is realized and hence self-modifying programs are possible—was only invented in the 1940s. Nevertheless, the history of computing (interpreted as the usage of modern computers) is only understandable against the background of the many forms of information processing as well as mechanical computing devices that solved mathematical problems in the first half of the twentieth century. The part these several predecessors played in the invention and early history of the computer may be interpreted from two different perspectives: on the one hand it can be argued that these machines prepared the way for the modern digital computer, on the other hand it can be argued that the computer, which was invented as a mathematical instrument, was reconstructed to be a data-processing machine, a control mechanism, and a communication tool.

The invention and early history of the digital computer has its roots in two different kinds of developments: first, information processing in business and government bureaucracies (see Calculators, Mechanical and Electromechanical); and second, the use and the search for mathematical instruments and methods that could solve mathematical problems arising in the sciences and in engineering (see Computers, Early Digital).

Origins in Mechanical Office Equipment

The development of information processing in business and government bureaucracies had its origins in the late nineteenth century, which was not just an era of industrialization and mass production but also a time of continuous growth in administrative work. The economic precondition for this development was the creation of a global economy, which caused growth in production of goods and trade. This brought with it an immense increase in correspondence, as well as monitoring and accounting activities—corporate bureaucracies began to collect and process data in increasing quantities. Almost at the same time, government organizations became more and more interested in collating data on population and demographic changes (e.g., expanding tax revenues, social security, and wide-ranging planning and monitoring functions) and analyzing this data statistically.

Bureaucracies in the U.S. and in Europe reacted in a different way to these changes. While in Europe for the most part neither office machines nor telephones entered offices until 1900, in the U.S. in the last quarter of the nineteenth century the information-handling techniques in bureaucracies were radically changed because of the introduction of mechanical devices for writing, copying, and counting data. The rise of big business in the U.S. had caused a growing demand for management control tools, which was fulfilled by a new ideology of systematic management together with the products of the rising office machines industry. Because of a later start in industrialization, the government and businesses in the U.S. were not forced to reorganize their bureaucracies when they introduced office machines. This, together with an ideological preference for modern office equipment, was the cause of a market for office machines and of a far-reaching mechanization of office work in the

U.S. In the 1880s typewriters and cash registers became very widespread, followed by adding machines and book-keeping machines in the 1890s. From 1880 onward, the makers of office machines in the U.S. underwent a period of enormous growth, and in 1920 the office machine industry annually generated about $200 million in revenue. In Europe, by comparison, mechanization of office work emerged about two decades later than in the U.S.—both Germany and Britain adopted the American system of office organization and extensive use of office machines for the most part no earlier than the 1920s.

During the same period the rise of a new office machine technology began. Punched card systems, initially invented by Herman Hollerith to analyze the U.S. census in 1890, were introduced. By 1911 Hollerith's company had only about 100 customers, but after it had been merged in the same year with two other companies to become the Computing-Tabulating-Recording Company (CTR), it began a tremendous ascent to become the world leader in the office machine industry. CTR's general manager, Thomas J. Watson, understood the extraordinary potential of these punched-card accounting devices, which enabled their users to process enormous amounts of data largely automatically, in a rapid way and at an adequate level of cost and effort. Due to Watson's insights and his extraordinary management abilities, the company (which had since been renamed to International Business Machines (IBM)) became the fourth largest office machine supplier in the world by 1928—topped only by Remington Rand, National Cash Register (NCR), and the Burroughs Adding Machine Company.

Origin of Calculating Devices and Analog Instruments

Compared with the fundamental changes in the world of corporate and government bureaucracies caused by office machinery during the late nineteenth and early twentieth century, calculating machines and instruments seemed to have only a minor influence in the world of science and engineering. Scientists and engineers had always been confronted with mathematical problems and had over the centuries developed techniques such as mathematical tables. However, many new mathematical instruments emerged in the nineteenth century and increasingly began to change the world of science and engineering. Apart from the slide rule, which came into popular use in Europe from the early nineteenth century onwards (and became the symbol of the engineer for decades), calculating machines and instruments were only produced on a large scale in the middle of the nineteenth century.

In the 1850s the production of calculating machines as well as that of planimeters (used to measure the area of closed curves, a typical problem in land surveying) started on different scales. Worldwide, less than 2,000 calculating machines were produced before 1880, but more than 10,000 planimeters were produced by the early 1880s. Also, various types of specialized mathematical analog instruments were produced on a very small scale in the late nineteenth century; among them were integraphs for the graphical solution of special types of differential equations, harmonic analyzers for the determination of Fourier coefficients of a periodic function, and tide predictors that could calculate the time and height of the ebb and flood tides (see Computers, Analog).

Nonetheless, in 1900 only geodesists and astronomers (as well as part of the engineering community) made extensive use of mathematical instruments. In addition, the establishment of applied mathematics as a new discipline took place at German universities on a small scale and the use of apparatus and machines as well as graphical and numerical methods began to flourish during this time. After World War I, the development of engineering sciences and of technical physics gave a tremendous boost to applied mathematics in Germany and Britain. In general, scientists and engineers became more aware of the capabilities of calculating machines and a change of the calculating culture—from the use of tables to the use of calculating machines—took place.

One particular problem that was increasingly encountered by mechanical and electrical engineers in the 1920s was the solution of several types of differential equations, which were not solvable by analytic solutions. As one important result of this development, a new type of analog instrument—the so called "differential analyzer"—was invented in 1931 by the engineer Vannevar Bush at the Massachusetts Institute of Technology (MIT). In contrast to its predecessors—several types of integraphs—this machine (which was later called an analog computer) could be used not only to solve a special class of differential equation, but a more general class of differential equations associated with engineering problems. Before the digital computer was invented in the 1940s there was an intensive use of analog instruments (similar to Bush's differential analyzer) and a number of

machines were constructed in the U.S. and in Europe after the model of Bush's machine before and during World War II. Analog instruments also became increasingly important in several fields such as the firing control of artillery on warships or the control of rockets (see Computers, Analog). It is worth mentioning here that only for a limited class of scientific and engineering problems was it possible to construct an analog computer—weather forecasting and the problem of shock waves produced by an atomic bomb, for example, required the solution of partial differential equations, for which a digital computer was needed.

The Invention of the Computer

The invention of the electronic digital stored-program computer is directly connected with the development of numerical calculation tools for the solution of mathematical problems in the sciences and in engineering. The ideas that led to the invention of the computer were developed simultaneously by scientists and engineers in Germany, Britain, and the U.S. in the 1930s and 1940s. The first freely programmable program-controlled automatic calculator was developed by the civil engineering student Konrad Zuse in Germany. Zuse started development work on program-controlled computing machines in the 1930s, when he had to deal with extensive calculations in static, and in 1941 his Z3, which was based on electromechanical relay technology, became operational.

Several similar developments in the U.S. were in progress at the same time. In 1937 Howard Aiken, a physics student at Harvard University, approached IBM to build a program-controlled calculator—later called the "Harvard Mark I." On the basis of a concept Aiken had developed because of his experiences with the numerical solution of partial differential equations, the machine was built and became operational in 1944. At almost the same time a series of important relay computers was built at the Bell Laboratories in New York following a suggestion by George R. Stibitz. All these developments in the U.S. were spurred by the outbreak of World War II. The first large-scale programmable electronic computer called the Colossus was built in complete secrecy in 1943 to 1944 at Bletchley Park in Britain in order to help break the German Enigma machine ciphers.

However, it was neither these relay calculators nor the Colossus that were decisive for the development of the universal computer, but the ENIAC (electronic numerical integrator and computer), which was developed at the Moore School of Engineering at the University of Pennsylvania. Extensive ballistic calculations were carried out there for the U.S. Army during World War II with the aid of the Bush "differential analyzer" and more than a hundred women ("computors") working on mechanical desk calculators. Observing that capacity was barely sufficient to compute the artillery firing tables, the physicist John W. Mauchly and the electronic engineer John Presper Eckert started developing the ENIAC, a digital version of the differential analyzer, in 1943 with funding from the U.S. Army.

In 1944 the mathematician John von Neumann turned his attention to the ENIAC because of his mathematical work on the Manhattan Project (on the implosion of the hydrogen bomb). While the ENIAC was being built, Neumann and the ENIAC team drew up plans for a successor to the ENIAC in order to improve the shortcomings of the ENIAC concept, such as the very small memory and the time-consuming reprogramming (actually rewiring) required to change the setup for a new calculation. In these meetings the idea of a stored-program, universal machine evolved. Memory was to be used to store the program in addition to data. This would enable the machine to execute conditional branches and change the flow of the program. The concept of a computer in the modern sense of the word was born and in 1945 von Neumann wrote the important "First Draft of a Report on the EDVAC," which described the stored-program, universal computer. The logical structure that was presented in this draft report is now referred to as the "von Neumann architecture." This EDVAC report was originally intended for internal use but once made freely available it became the "bible" for computer pioneers throughout the world in the 1940s and 1950s. The first computer featuring the von Neumann architecture operated at Cambridge University in the U.K.; in June 1949 the EDSAC (electronic delay storage automatic computer) computer built by Maurice Wilkes—designed according to the EDVAC principles—became operational (see Computers, Early Digital).

The Computer as a Scientific Instrument

As soon as the computer was invented, a growing demand for computers by scientists and engineers evolved, and numerous American and European universities started their own computer projects in the 1940s and 1950s. After the technical difficulties of building an electronic computer were solved, scientists grasped the opportunity to use the new

scientific instrument for their research. For example, at the University of Göttingen in Germany, the early computers were used for the initial value problems of partial differential equations associated with hydrodynamic problems from atomic physics and aerodynamics. Another striking example was the application of von Neumann's computer at the Institute for Advanced Study (IAS) in Princeton to numerical weather forecasts in 1950. As a result, numerical weather forecasts could be made on a regular basis from the mid-1950s onwards.

Mathematical methods have always been of a certain importance for science and engineering sciences, but only the use of the electronic digital computer (as an enabling technology) made it possible to broaden the application of mathematical methods to such a degree that research in science, medicine, and engineering without computer-based mathematical methods has become virtually inconceivable at the end of the twentieth century. A number of additional computer-based techniques, such as scientific visualization, medical imaging, computerized tomography, pattern recognition, image processing, and statistical applications, have become of the utmost significance for science, medicine, engineering, and social sciences. In addition, the computer changed the way engineers construct technical artifacts fundamentally because of the use of computer-based methods such as computer-aided design (CAD), computer-aided manufacture (CAM), computer-aided engineering, control applications, and finite-element methods (see Computer-Aided Design and Manufacture). However, the most striking example seems to be the development of scientific computing and computer modeling, which became accepted as a third mode of scientific research that complements experimentation and theoretical analysis (see Computer Modeling). Scientific computing and computer modeling are based on supercomputers as the enabling technology, which became important tools for modern science routinely used to simulate physical and chemical phenomena. These high-speed computers became equated with the machines developed by Seymour Cray, who built the fastest computers in the world for many years. The supercomputers he launched such as the legendary CRAY I from 1976 were the basis for computer modeling of real world systems, and helped, for example, the defense industry in the U.S. to build weapons systems and the oil industry to create geological models that show potential oil deposits (see Computers, Supercomputers).

Growth of Digital Computers in Business and Information Processing

When the digital computer was invented as a mathematical instrument in the 1940s, it could not have been foreseen that this new artifact would ever be of a certain importance in the business world. About 50 firms entered the computer business worldwide in the late 1940s and the early 1950s, and the computer was reconstructed to be a type of electronic data-processing machine that took the place of punched-card technology as well as other office machine technology. It is interesting to consider that there were mainly three types of companies building computers in the 1950s and 1960s: newly created computer firms (such as the company founded by the ENIAC inventors Eckert and Mauchly), electronics and control equipments firms (such as RCA and General Electric), and office appliance companies (such as Burroughs and NCR). Despite the fact that the first digital computers were put on the market by a German and a British company, U.S. firms dominated the world market from the 1950s onward, as these firms had the biggest market as well as financial support from the government.

Generally speaking, the Cold War exerted an enormous influence on the development of computer technology. Until the early 1960s the U.S. military and the defense industry were the central drivers of the digital computer expansion, serving as the main market for computer technology and shaping and speeding up the formation of the rising computer industry. Because of the U.S. military's role as the "tester" for prototype hard- and software, it had a direct and lasting influence on technological developments; in addition, it has to be noted that the spread of computer technology was partly hindered by military secrecy. Even after the emergence of a large civilian computer market in the 1960s, the U.S. military maintained its influence by investing a great deal in computer in hard- and software and in computer research projects.

From the middle of the 1950s onwards the world computer market was dominated by IBM, which accounted for more than 70 percent of the computer industry revenues until the mid-1970s. The reasons for IBM's overwhelming success were diverse, but the company had a unique combination of technical and organizational capabilities at its disposal that prepared it perfectly for the mainframe computer market. In addition, IBM benefited from enormous government contracts, which helped to develop excellence in computer technology and design. However, the greatest advantage of

IBM was by no doubt its marketing organization and its reputation as a service-oriented firm, which was used to working closely with customers to adapt machinery to address specific problems, and this key difference between IBM and its competitors persisted right into the computer age.

During the late 1950s and early 1960s, the computer market—consisting of IBM and seven other companies called the "seven dwarves"—was dominated by IBM, with its 650 and 1401 computers. By 1960 the market for computers was still small. Only about 7,000 computers had been delivered by the computer industry, and at this time even IBM was primarily a punched-card machine supplier, which was still the major source of its income. Only in 1960 did a boom in demand for computers start, and by 1970 the number of computers installed worldwide had increased to more than 100,000. The computer industry was on the track to become one of the world's major industries, and was totally dominated by IBM.

The outstanding computer system of this period was IBM's System/360. It was announced in 1964 as a compatible family of the same computer architecture, and employed interchangeable peripheral devices in order to solve IBM's problems with a hotchpotch of incompatible product lines (which had evoked large problems in the development and maintenance of a great deal of different hardware and software products). Despite the fact that neither the technology used nor the systems programming were of a high-tech technology at the time, the System/360 established a new standard for mainframe computers for decades. Various computer firms in the U.S., Europe, Japan and even Russia, concentrated on copying components, peripherals for System/360 or tried to build System/360-compatible computers.

The growth of the computer market during the 1960s was accompanied by market shakeouts: two of the "seven dwarves" left the computer business after the first computer recession in the early 1970s, and afterwards the computer market was controlled by IBM and BUNCH (Burroughs, UNIVAC, NCR, Control Data, and Honeywell). At the same time, an internationalization of the computer market took place—U.S. companies controlled the world market for computers—which caused considerable fears over loss of national independence in European and Japanese national governments, and these subsequently stirred up national computing programs. While the European attempts to create national champions as well as the more general attempt to create a European-wide market for mainframe computers failed in the end, Japan's attempt to found a national computer industry has been successful: Until today Japan is the only nation able to compete with the U.S. in a wide array of high-tech computer-related products (see Computers, Mainframe).

Real-Time and Time-Sharing

Until the 1960s almost all computers in government and business were running batch-processing applications (i.e., the computers were only used in the same way as the punched-card accounting machines they had replaced). In the early 1950s, however, the computer industry introduced a new mode of computing named "real-time" in the business sector for the first time, which was originally developed for military purposes in MIT's Whirlwind project. This project was initially started in World War II with the aim of designing an aircraft simulator by analog methods, and later became a part of a research and development program for the gigantic, computerized anti-aircraft defense system SAGE (semi-automatic ground environment) built up by IBM in the 1950s.

The demand for this new mode of computing was created by cultural and structural changes in economy. The increasing number of financial transactions in banks and insurance companies as well as increasing airline traveling activities made necessary new computer-based information systems that led finally to new forms of business evolution through information technology.

The case of the first computerized airline reservation system SABRE, developed for American Airlines by IBM in the 1950s and finally implemented in the early 1960s, serves to thoroughly illustrate these structural and structural changes in economy. Until the early 1950s, airline reservations had been made manually without any problems, but by 1953 this system was in crisis because increased air traffic and growing flight plan complexity had made reservation costs insupportable. SABRE became a complete success, demonstrating the potential of centralized real-time computing systems connected via a network. The system enabled flight agents throughout the U.S., who were equipped with desktop terminals, to gain a direct, real-time access to the central reservation system based on central IBM mainframe computers, while the airline was able to assign appropriate resources in response. Therefore an effective combination of advantages was offered by SABRE—a better utilization of resources and a much higher customer convenience.

Very soon this new mode of computing spread around the business and government world and became commonplace throughout the service and distribution sectors of the economy; for example, bank tellers and insurance account representatives increasingly worked at terminals. On the one hand structural information problems led managers to go this way, and on the other hand the increasing use of computers as information handling machines in government and business had brought about the idea of computer-based accessible data retrieval. In the end, more and more IBM customers wanted to link dozens of operators directly to central computers by using terminal keyboards and display screens.

In the late 1950s and early 1960s—at the same time that IBM and American Airlines had begun the development of the SABRE airline reservation system—a group of brilliant computer scientists had a new idea for computer usage named "time sharing." Instead of dedicating a multiterminal system solely to a single application, they had the computer utility vision of organizing a mainframe computer so that several users could interact with it simultaneously. This vision was to change the nature of computing profoundly, because computing was no longer provided to naïve users by programmers and systems analysts, and by the late 1960s time-sharing computers became widespread in the U.S.

Particularly important for this development had been the work of J.C.R. Licklider of the Advanced Research Project Agency (ARPA) of the U.S. Department of Defense. In 1960 Licklider had published a now-classic paper "Man–Computer Symbiosis" proposing the use of computers to augment human intellect and creating the vision of interactive computing. Licklider was very successful in translating his idea of a network allowing people on different computers to communicate into action, and convinced ARPA to start an enormous research program in 1962. Its budget surpassed that of all other sources of U.S. public research funding for computers combined. The ARPA research programs resulted in a series of fundamental moves forward in computer technology in areas such as computer graphics, artificial intelligence (see Artificial Intelligence), and operating systems. For example, even the most influential current operating system, the general-purpose time-sharing system Unix, developed in the early 1970s at the Bell Laboratories, was a spin-off of an ambitious operating system project, Multics, funded by ARPA. The designers of Unix successfully attempted to keep away from complexity by using a clear, minimalist design approach to software design, and created a multitasking, multiuser operating system, which became the standard operating system in the 1980s (see Computers, Systems Programs).

Electronic Component Revolution

While the nature of business computing was changed by the new paradigms such as real time and time sharing, advances in solid-state components increasingly became a driving force for fundamental changes in the computer industry, and led to a dynamic interplay between new computer designs and new programming techniques that resulted in a remarkable series of technical developments. The technical progress of the mainframe computer had always run parallel to conversions in the electronics components (see Computer Memory, Main). During the period from 1945 to 1965, two fundamental transformations in the electronics industry took place that were marked by the invention of the transistor in 1947 and the integrated circuit in 1957 to 1958. While the first generation of computers—lasting until about 1960—was characterized by vacuum tubes (valves) for switching elements, the second generation used the much smaller and more reliable transistors, which could be produced at a lower price. A new phase was inaugurated when an entire integrated circuit on a chip of silicon was produced in 1961, and when the first integrated circuits were produced for the military in 1962. A remarkable pace of progress in semiconductor innovations, known as the "revolution in miniature," began to speed up the computer industry. The third generation of computers characterized by the use of integrated circuits began with the announcement of the IBM System/360 in 1964 (although this computer system did not use true integrated circuits). The most important effect of the introduction of integrated circuits was not to strengthen the leading mainframe computer systems, but to destroy Grosch's Law, which stated that computing power increases as the square of its costs. In fact, the cost of computer power dramatically reduced during the next ten years.

This became clear with the introduction of the first computer to use integrated circuits on a full scale in 1965: the Digital Equipment Corporation (DEC) offered its PDP-8 computer for just $18,000, creating a new class of computers called minicomputers—small in size and low in cost—as well as opening up the market to new customers. Minicomputers were mainly used in areas other

than general-purpose computing such as industrial applications and interactive graphics systems. The PDP-8 became the first widely successful minicomputer with over 50,000 items sold, demonstrating that there was a market for smaller computers. This success of DEC (by 1970 it had become the world's third largest computer manufacturer) was supported by dramatic advances in solid-state technology. During the 1960s the number of transistors on a chip doubled every two years, and as a result minicomputers became continuously more powerful and more inexpensive at an inconceivable speed.

Personal Computing

The most striking aspect of the consequences of the exponential increase of the number of transistors on a chip during the 1960s—as stated by "Moore's Law": the number of transistors on a chip doubled every two years—was not the lowering of the costs of mainframe computer and minicomputer processing and storage, but the introduction of the first consumer products based on chip technology such as hand-held calculators and digital watches in about 1970 (see Calculators, Electronic). More specifically, the market acts in these industries were changed overnight by the shift from mechanical to chip technology, which led to an enormous deterioration in prices as well as a dramatic industry shakeout. These episodes only marked the beginning of wide-ranging changes in economy and society during the last quarter of the twentieth century leading to a new situation where chips played an essential role in almost every part of business and modern life.

The case of the invention of the personal computer serves to illustrate that it was not sufficient to develop the microprocessor as the enabling technology in order to create a new invention, but how much new technologies can be socially constructed by cultural factors and commercial interests. When the microprocessor, a single-chip integrated circuit implementation of a CPU, was launched by the semiconductor company Intel in 1971, there was no hindrance to producing a reasonably priced microcomputer, but it took six years until the consumer product PC emerged. None of the traditional mainframe and minicomputer companies were involved in creating the early personal computer. Instead, a group of computer hobbyists as well as the "computer liberation" movement in the U.S. became the driving force behind the invention of the PC. These two groups were desperately keen on a low-priced type of minicomputer for use at home for leisure activities such as computer games; or rather they had the counterculture vision of an unreservedly available and personal access to an inexpensive computer utility provided with rich information. When in 1975 the Altair 8800, an Intel 8080 microprocessor-based computer, was offered as an electronic hobbyist kit for less than $400, these two groups began to realize their vision of a "personal computer." Very soon dozens of computer clubs and computer magazines were founded around the U.S., and these computer enthusiasts created the personal computer by combining the Altair with keyboards, disk drives, and monitors as well as by developing standard software for it. Consequently in only two years, a more or less useless hobbyist kit had been changed into a computer that could easily be transformed in a consumer product.

The computer hobbyist period ended in 1977, when the first standard machines for an emerging consumer product mass market were sold. These included products such as the Commodore Pet and the Apple II, which included its own monitor, disk drive, and keyboard, and was provided with several basic software packages. Over next three years, spreadsheet, word processing, and database software were developed, and an immense market for games software evolved (see Computers, Software, Application Programs). As a result, personal computers became more and more a consumer product for ordinary people, and Apple's revenues shot to more than $500 million in 1982. By 1980, the personal computer had transformed into a business machine, and IBM decided to develop its own personal computer, which was introduced as the IBM PC in 1981. It became an overwhelming success and set a new industry standard.

Apple tried to compete by launching their new Macintosh computer in 1984 provided with a revolutionary graphical user interface (GUI), which set a new standard for a user-friendly human–computer interaction. It was based on technology created by computer scientists at the Xerox Palo Alto Research Center in California, who had picked up on ideas about human–computer interaction developed at the Stanford Research Institute and at the University of Utah. Despite the fact that the Macintosh's GUI was far superior to the MS-DOS operating system of the IBM-compatible PCs, Apple failed to win the business market and remained a niche player with a market share of about 10 percent. The PC main branch was determined by the companies IBM had

chosen as its original suppliers in 1981 for the design of the microprocessor (Intel) and the operating system (Microsoft). While IBM failed to seize power in the operating system software market for PCs in a software war with Microsoft, Microsoft achieved dominance not only of the key market for PC operating systems, but also the key market of office applications during the first half of the 1990s (see Computers, Personal).

Networking

In the early 1990s computing again underwent further fundamental changes with the appearance of the Internet, and for the most computer users, networking became an integral part of what it means to have a computer (see Computer Networks). Furthermore, the rise of the Internet indicated the impending arrival of a new "information infrastructure" as well as of a "digital convergence," as the coupling of computers and communications networks was often called.

In addition the 1990s were a period of an information technology boom, which was mainly based on the Internet hype. For many years previously, it seemed to a great deal of managers and journalists that the Internet would become not just an indispensable business tool, but also a miracle cure for economic growth and prosperity. In addition, computer scientists and sociologists started a discussion predicting the beginning of a new "information age" based on the Internet as a "technological revolution" and reshaping the "material basis" of industrial societies (see Internet).

The Internet was the outcome of an unusual collaboration of a military–industrial–academic complex that promoted the development of this extraordinary innovation. It grew out of a military network called the ARPAnet, a project established and funded by ARPA in the 1960s. The ARPAnet was initially devoted to support of data communications for defense research projects and was only used by a small number of researchers in the 1970s. Its further development was primarily promoted by unintentional forms of network usage. The users of the ARPAnet became very much attracted by the opportunity for communicating through electronic mail, which rapidly surpassed all other forms of network activities. Another unplanned spin-off of the ARPAnet was the Usenet (Unix User Network), which started in 1979 as a link between two universities and enabled its users to subscribe to newsgroups. Electronic mail became a driving force for the creation of a large number of new proprietary networks funded by the existing computer services industry or by organizations such as the NSF (NSFnet). Because networks users' desire for email to be able to cross network boundaries, an ARPA project on "internetworking" became the origin for the "Internet"—a network of networks linked by several layers of protocols such as TCP/IP (transmission control protocol/internet protocol), which quickly developed into the actual standard.

Only after the government funding had solved many of the most essential technical issues and had shaped a number of the most characteristic features of the Internet, did private sector entrepreneurs start Internet-related ventures and quickly developed user-oriented enhancements. Nevertheless the Internet did not make a promising start and it took more than ten years before significant numbers of networks were connected. In 1980, the Internet had less than two hundred hosts, and during the next four years the number of hosts went up only to 1000. Only when the Internet reached the educational and business community of PC users in the late 1980s, did it start to become an important economic and social phenomenon. The number of hosts began an explosive growth in the late 1980s—by 1988 there were over 50,000 hosts. An important and unforeseen side effect of this development became the creation of the Internet into a new electronic publishing medium. The electronic publishing development that excited most interest in the Internet was the World Wide Web, originally developed at the CERN High Energy Physics Laboratory in Geneva in 1989 (see World Wide Web). Soon there were millions of documents on the Internet, and private PC users became excited by the joys of surfing the Internet. A number of firms such as AOL soon provided low-cost network access and a range of consumer-oriented information services. The Internet boom was also helped by the Clinton–Gore presidential election campaign on the "information superhighway" and by the amazing news reporting on the national information infrastructure in the early 1990s. Nevertheless for many observers it was astounding how fast the number of hosts on the Internet increased during the next few years—from more than 1 million in 1992 to 72 million in 1999.

The overwhelming success of the PC and of the Internet tends to hide the fact that its arrival marked only a branching in computer history and not a sequence. (Take, for example, the case of mainframe computers, which still continue to run, being of great importance to government facilities and the private sector (such as banks and insurance

companies), or the case of supercomputers, being of the utmost significance for modern science and engineering.) Furthermore it should be noted that only a small part of the computer applications performed today is easily observable—98 percent of programmable CPUs are used in embedded systems such as automobiles, medical devices, washing machines and mobile telephones.

See also **Computer and Video Games; Computer Displays; Computer Memory; Computer Science; Computers, Hybrid; Computer–User Interface; Software Engineering**

ULF HASHAGEN

Further Reading

Abbate, J. *Inventing the Internet*. Cambridge, MA, 1999.
Aspray, W. *John von Neumann and the Origins of Modern Computing*. Cambridge, MA, 1990.
Campbell-Kelly, M. and Aspray, W. *Computer: A History of the Information Machine*. New York, 1996.
Campbell-Kelly, M. *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry*. Cambridge, MA, 2003.
Ceruzzi, P. *A History of Modern Computing*. Cambridge, MA, 1998.
Cortada, J. *Before the Computer: IBM, NCR, Burroughs, and Remington Rand and the Industry They Created, 1856–1956*. Princeton, NJ, 1993.
Edwards, P. *The Closed World: Computers and the Politics of Discourse in Cold War America*. Cambridge, MA, 1996.
Hashagen, U., Keil-Slawik and Norberg, A., Eds. *History of Computing: Software Issues*. Berlin, 2002.
MacKenzie, D. *Mechanizing Proof: Computing, Risk, and Trust*. Cambridge, MA, 2001.
Norberg, A. and O'Neill, J. *Transforming Computer Technology: Information Processing for the Pentagon, 1962-1986*. Baltimore, 1996.
Petzold, H. *Rechnende Maschinen: eine historische Untersuchung ihrer Herstellung und Anwendung vom Kaiserreich bis zur Bundesrepublik*. Düsseldorf, 1985.
Roland, A. *DARPA and the Quest for Machine Intelligence, 1983–1993*. Cambridge, MA, 2002.

Computer–User Interface

A computer interface is the point of contact between a person and an electronic computer. Today's interfaces include a keyboard, mouse, and display screen. Computer user interfaces developed through three distinct stages, which can be identified as batch processing, interactive computing, and the graphical user interface.

Early computing machines used punch cards as a method to input instructions to be calculated. Punch-card technology was used during the first half of the twentieth century for business machines and it was applied to early computers. The ENIAC (electronic numerical integrator and computer), developed between 1943 and 1945 at the Moore School of Electrical Engineering at the University of Pennsylvania, used punch card technology and was the first electronic computer. Human operators supplied instructions to the ENIAC on punched cards and then waited for the results. For the following two decades, punch cards were the main input devices for computer–user interfaces.

In the postwar era, new types of interactive computer user interfaces were invented. The technologies required to enable people to interactively operate a computer were first developed in American government-funded projects, such as the SAGE (semi-automatic ground environment) air defense system, based on MIT's Whirlwind Computer. Project SAGE introduced magnetic core memory as a method of information storage to replace punched cards. Additionally, SAGE designers wrote software to run the computer's CRT (cathode-ray tube) that displayed information to human operators. The system converted radar information into pictures generated on the CRT and operators could select information by pointing to a computer-generated target with a light pen. This type of interactive interface was further advanced by Ivan Sutherland's Sketchpad project.

In 1962, Sketchpad introduced the concept of computer graphics by enabling an operator to create sophisticated visual models on a display screen that resembled a television set. Operators created models by putting information into the computer with a light pen and keyboard. Computer-generated drawings were immediately updated on the screen as different commands were selected.

Sketchpad was part of a new movement that examined how users interacted with computers. "Human–computer interaction" (HCI) is a term used to describe the psychology of how people interact and use computers. HCI combines computer science, psychology and ergonomics to create software to be used as part of the computer–user interface. The study of HCI has led to the development of computer systems that are easier to use. Ergonomics, or human factors, is the study of applying psychology to interface design to make computers easier to operate and understand. As computers became more interactive, the ergonomic features of interface design advanced to the point at which average people could easily operate them.

A pioneer in developing interactive computing was Douglas C. Engelbart. While working at the Stanford Research Laboratory in the 1960s, Engelbart created many of the tools that are

commonly used with today's personal computers, including the mouse, graphical interfaces, hypertext, multiple windows, and teleconferencing. Engelbart's system, the "augment system," replaced the light pen with a mouse. The original mouse was a small input device in the shape of a box with hidden wheels. As the mouse rolled on a flat surface the wheels signaled a cursor or symbol displayed on the computer screen.

Engelbart's augment system interactively displayed text, graphics, and video images. Computer commands were input through a keyboard, mouse, and five-finger keyset. In December 1968, Engelbart and his team demonstrated his vision of interactive computing at the Association for Computer Machinery/Institute of Electrical Engineers Fall Joint Computer Conference in San Francisco. Sitting in the audience was a young computer scientist named Alan Kay.

In the early 1970s, Kay (along with members of Engelbart's team) joined the newly established Xerox Palo Alto Research Center (PARC). The PARC researchers developed the hardware and software necessary for the next generation of computer–user interfaces, which included pictures, windows, and menus.

At Xerox PARC, researchers designed and built a new computer system called the Alto, which had enough power to drive a full-screen graphical image. The Alto used bitmap technology that created a one-to-one correspondence between the picture elements on the screen and the bits in the computer's memory. Bitmapping enables users to scale letters and mix text and graphics together on a display screen. The Alto's text-editing software added a new feature called "what you see is what you get" (WYSIWYG). Images displayed on the screen visually resembled the computer-generated information output on a printed page. These new technological developments enabled PARC researchers to invent the WIMP interface. WIMP interfaces incorporate windows, icons, a mouse, and pull-down menus into a visual interface design.

Kay and his team further enhanced the WIMP design by transforming it into the graphical user interface (GUI). GUIs enable computer users to execute commands by pointing and clicking on icons. The icons displayed on the screen resemble familiar office objects such as file folders, file cabinets, and in-mail and out-mail baskets. Objects displayed on the screen are manipulated with a mouse and the screen's work area metaphorically represents a desktop. PARC's GUI added the desktop metaphor to computer–user interfaces.

In 1979, after seeing a demonstration of the Xerox PARC technology, Steve Jobs saw the possibility of using graphical interfaces to make computers more "user-friendly." He applied PARC technology to the Macintosh and it became the first commercially successful GUI. In 1984, Apple advertised the Macintosh as a "user-friendly" system that easily enabled average people to operate a computer. The Macintosh also popularized the desktop metaphor developed at PARC. Following the success of the Macintosh interface, a number of different companies introduced graphical and WIMP-styled interfaces, including International Business Machines (IBM), Digital Research, and Commodore. However, the most successful GUI was Microsoft Windows, which still dominates the interface marketplace. Currently, GUIs have become the most popular method of computer–user interaction.

Today's graphical interfaces support additional multimedia features, such as streaming audio and video. In GUI design, every new software feature introduces more icons into the process of computer–user interaction. Presently, the large vocabulary of icons used in GUI design is difficult for users to remember, which creates a complexity problem. As GUIs become more complex, interface designers are adding voice recognition and intelligent agent technologies to make computer user interfaces even easier to operate.

See also **Computer Displays**

SUSAN B. BARNES

Further Reading

Barnes, S.B. Engelbart, D.C. Developing the underlying concepts for contemporary computing. *IEEE Ann. Hist. Comput.*, 19, 3, 16–26, 1997.

Campbell-Kelly, M. and Aspray, W. *Computer: A History of the Information Machine*. Harper Collins, New York, 1996.

Freiberger, P. and Swaine, M. *Fire in the Valley: The Making of the Personal Computer*, 2nd edn. McGraw-Hill, New York, 2000.

Hiltzik, M.A. *Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age*. Harper Collins, New York, 1999.

Kay, A.C. *User Interface: A Personal View, in The Art of Human-Computer Interface Design*, Laurel, B., Ed. Addison-Wesley, Reading, MA, 1990.

Kay, A. *The Early History of Smalltalk in Programming Languages*. Bergin, T.J. and Gibson, R.G. ACM Press, New York, 1996.

Rheingold, H. *Tools for Thought*. Simon & Schuster, New York, 1985.

Rose, F. *West of Eden: The End of Innocence at Apple Computer*. Penguin, New York, 1989.

Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd edn. Addison-Wesley, Boston, 1997.
Smith, D.K. and Alexander, R.C. *Fumbling the Future: How Xerox Invented, then Ignored, the First Personal Computer*. William Morrow, New York, 1988.

Concrete, Reinforced

Reinforced concrete was in its infancy at the opening of the twentieth century, but it was very quickly adopted worldwide as an economic and versatile construction material. Employing fairly basic materials—sand, crushed stone or gravel, cement, and steel—it found use in all the existing aspects of construction, including buildings, roads, bridges, dams, reservoirs, and docks. It also served the century's new applications, such as air raid shelters and the pressure vessels of nuclear reactors. By the end of the twentieth century, concrete in its various forms—plain, reinforced, and prestressed—was probably the most widely used construction material in the world.

Although concrete had been used from at least Roman times, it was not until the last decades of the nineteenth century that the idea of reinforcing it was applied to construction. Until then, concrete was used as a cheaper and more versatile substitute for stone and brick masonry, with which it shared the properties of being strong in compression but weak and brittle when subject to tension or bending. Not until rods or bars of wrought iron, and later steel, were embedded in concrete was it considered worthwhile to use the resulting reinforced concrete with confidence for floor and roof slabs, beams, trusses, cantilevers, and other elements that work by bending.

Proprietary reinforced concrete systems were patented and used almost simultaneously in several countries from the 1880s, notably the U.S., France, and Germany, followed quickly by Britain, where the Frenchman François Hennebique had taken out a patent in 1892. His system used the plain round mild steel bar, but other systems adopted alternative profiles (Figure 14) both to satisfy the need for originality if a patent were to be granted, and to ensure good grip or bond with the concrete. This was essential if the concrete and its reinforcement were to work together. The other key component of modern concrete—Portland cement, patented in 1824 by Joseph Aspdin—had by this time been developed as a strong and reliable product.

Guidance on the use, design, and construction of reinforced concrete quickly became available. The first textbook in the U.K., by Marsh, appeared

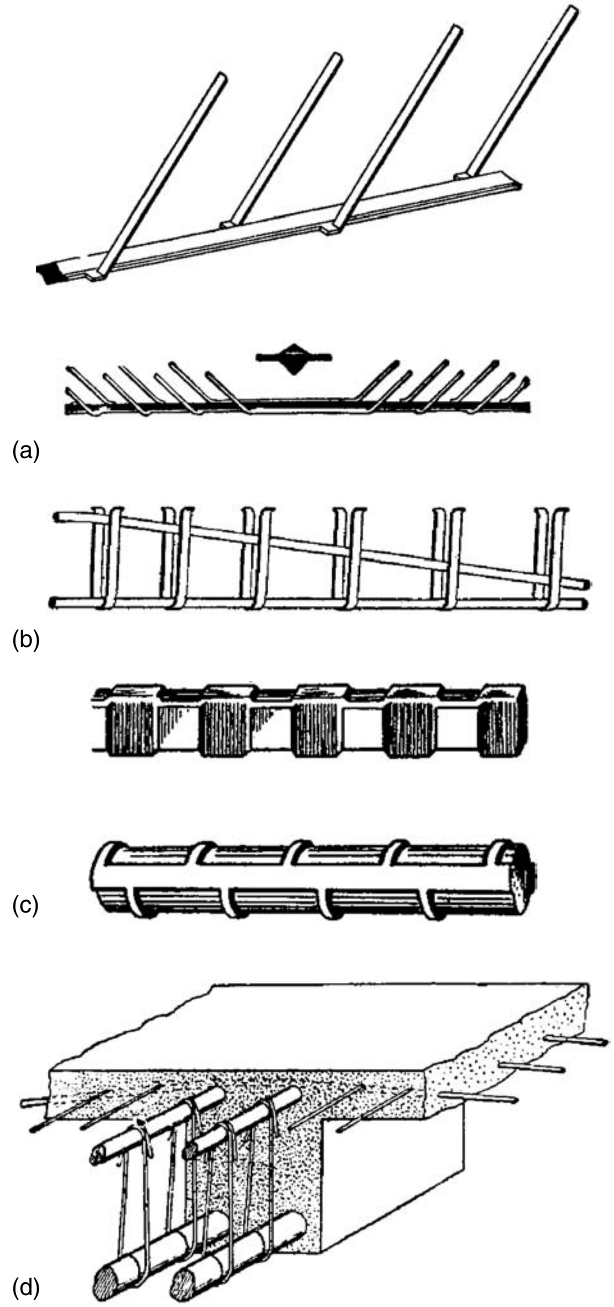


Figure 14. Some common early twentieth century reinforcing systems: (a) Coignet; (b) Hennebique; (c) Indented bar; and (d) Kahn.
[*Source: Jones, B.E., Ed. Cassell's Reinforced Concrete. Waverley Book Company, London, 1920*.]

in 1904. Subsequently, codes of practice for design and standards for materials were introduced. Subsequently, as later, research and development made major contributions to the understanding of concrete behavior through materials testing and load tests that studied the behavior of concrete structures. Many committees, learned societies,

and trade associations were established during the century, providing the means of exchanging experience and understanding, both nationally and internationally.

As patents expired during the earlier decades of the century, reinforced concrete construction became freely available to contractors. This was a mixed blessing, as some lacked the skills and experience of the system providers, who took pains to ensure that their products were soundly designed and built.

A notable development was discovered by Eugène Freyssinet in 1928. He pioneered prestressing, using tensioned steel to precompress the whole of the concrete section, so that when loaded in service it did not develop tension stresses. The result was a more efficient member with improved durability.

The major reconstruction program undertaken in many countries in the decades after World War II made very extensive use of concrete. Inexperience, and pressures of time and money, resulted in many cases of inadequate cover of reinforcement and other defects so that subsequently, major repairs were needed to remedy the consequences of corroding reinforcement. By the end of the century, however, the issues of durability and how to achieve it were widely understood, so that any competent engineer or contractor could now be expected to build soundly in concrete.

Much research, development, and innovation was applied during the century to the essential materials used in concrete—cement, aggregates, and reinforcement.

Cement

Cements were developed for particular applications, such as rapid setting or resisting aggressive environments such as seawater. Others made use of industrial waste products such as ground granulated blastfurnace slag and fly ash (the ground-up clinker from coal-burning power stations). Inevitably, some materials proved to have unwelcome side effects, such as the very quick-setting high alumina cement, which was found in the 1970s to lose strength over time, even at room temperature. This was originally thought to occur only in warm damp environments. Similarly, calcium chloride, commonly added to the concrete mix to accelerate setting, especially in cold weather, was found subsequently to increase the risk of reinforcement corrosion. Use of both materials is now restricted by codes of practice.

Aggregates

The most common aggregates remain sand and crushed rock or gravel. However, commercial incentives and, later in the twentieth century, environmental issues such as opposition to new quarrying for natural aggregates, led to alternatives being sought and developed. Lightweight aggregates were made cheaply from waste materials—initially clinker from coal-burning and broken bricks, and later (once again) fly ash. Fly ash, when heated, forms pellets suitable for use as aggregate. Lightweight concrete offers savings in the supporting structure, and also offers better thermal insulation than normal-weight concrete. Concern over energy conservation issues in buildings from the 1970s meant that lightweight concrete blocks have become very widely used for the inner leaf of cavity wall construction, meeting onerous building regulation requirements for thermal insulation.

Reinforcement

With improved steel-making techniques and the need to ensure good bond to the concrete, the ribbed hot-rolled high-yield strength steel bar had displaced the plain mild steel bar to become the norm in the U.K., Europe, and elsewhere by the end of the century. Higher strength steel was needed for prestressed concrete. Forms developed and still in use included rods that could be pretensioned against the molds for precast unit manufacture, and cables or threaded bars that could be post-tensioned inside sheaths cast into the concrete, the latter approach often being used for larger beams, especially in bridges, cast *in situ*.

Concerns over durability led to the use of galvanized and epoxy-coated or stainless steel reinforcement, particularly on bridges and in car parks where deicing salt carried by vehicles can soak into concrete and accelerate the corrosion process. Although more expensive than plain steel, the greater initial cost may be outweighed by the potential savings from reduced—and necessarily disruptive—future maintenance and repair costs. Similar arguments apply to carbon fiber polymer-based reinforcement, which was still in its infancy at the turn of the twenty-first century.

Long-established and widely used throughout the twentieth century were asbestos sheet, corrugated asbestos, and woodwool, whose names belie the fact that all three are early forms of fiber-reinforced cement. Flat asbestos-cement sheet originated in Austria around 1900, while the corrugated form—which could span longer distances, and so was ideally suited for pitched roofs

on factories and sheds—was being made in Britain by 1914. Woodwool is also believed to have originated at about this time in Austria, making use of waste timber shavings bound together with cement. Pressed into slabs, the woodwool provides lightweight roofing and walling panels with good thermal insulation properties.

Steel and polypropylene fibers have been used to reinforce concrete, particularly ground-bearing slabs. Their advocates argue that they reduce the incidence of cracking, although care is needed to obtain an even distribution of the fibers throughout the concrete mix. Glass fiber, of very light weight and capable of being molded into esthetically pleasing curves, has found use in thin glass-reinforced cement (GRC) cladding panels.

See also **Bridges, Concrete; Concrete Shells; Dams**

MICHAEL BUSSELL

Further Reading

Hamilton, S.B. *A Note on the History of Reinforced Concrete in Buildings: National Building Studies Special Report No. 24*. Her Majesty's Stationery Office, London, 1956.

Jones, B.E., Ed. *Cassell's Reinforced Concrete*. Waverley Book Company, London, 1920.

Mainstone, R.J. *Developments in Structural Form*. Architectural Press, Oxford, 1998.

Marsh, C.F. *Reinforced Concrete*. Constable, London, 1904.

Neville, A.M. *Properties of Concrete*. Longman, Harlow; and Wiley, New York, 1981.

Sutherland, J., Humm, D. and Chrimes, M., Eds. *Historic Concrete: Background to Appraisal*. Thomas Telford, London, 2001.

Concrete Shells

Of all the developments in the structural engineering of buildings in the last century, the concrete shell was surely the most spectacular. It provided the means of covering vast areas with a shell of reinforced concrete just a few centimeters thick. Like most developments in building engineering, the origins of shell structures have many strands. Roman engineers constructed domes and barrel vault roofs made of brick or concrete spanning of up to 40 meters, but these were relatively thick—over a meter at their thinnest part. In Gothic cathedrals, at up to 20 meters, spans were more modest but they were often much thinner—as little as 200 millimeters. There were also vernacular precedents, most prominently the thin tile vaults widely used in Catalonia from the seventeenth century which, made using quick-setting gypsum mortar, had the advantage that they could be built without the need for a supporting structure during construction. The idea was exported to the U.S. and patented in the late nineteenth century by Guastavino who used them in many hundreds of buildings, including a spectacular roof at the Pennsylvania Railway station.

The characteristic of a shell structure is the very high ratio of span to thickness. While a concrete beam or floor might achieve a ratio of 20 or perhaps thirty to one, a shell structure can achieve ratios of many hundreds to one. Eduardo Torroja's market hall (1933) at Algeciras in Spain is some 48 meters in diameter, yet is just 90 millimeters thick; and his roof at Madrid race course (1935) cantilevers 13 meters with a shell just 50 millimeters thick. (Figure 15).

The "ideal" concrete shell is a curved sheet of solid material carrying its loads in the plane of the shell since it is too thin to resist bending. Since a roof must be able to carry a variety of different loads caused by wind or snow, real concrete shell roofs do need to have some ability to resist bending. A shell curved in just one direction, like a sheet of paper curved to form a barrel vault, is far too flexible to be useful and most concrete shells gain their rigidity by being curved in two directions, for instance like an egg shell. Such a shell can carry normal loads by arching, which develops reactive tension forces in the plane of the shell. It thus effectively carries its loads as in-plane shear forces. The greater the curvature of a doubly curved shell, the greater its ability to resist bending. However, it is also a requirement of most roofs that they be not too curved; and there is also the need to ensure the thin shell surface does not buckle in compression. There is, then, a need to stiffen the shell itself and this has generally been achieved in three ways—thickening and reinforcing the shell, folding or corrugating the shell, or providing stiffening ribs. In fact, nature had developed all these approaches in the structures of plants, seashells and, of course, the egg shell (which has a span-to-thickness ratio of about 130). Finally, of course, the reinforced concrete shell needs to have sufficient thickness for two layers of reinforcement, roughly orthogonal, and sufficient concrete cover to keep water out.

The earliest concrete shells were developed by reinforced concrete engineers and contractors who, especially from the 1910s to the 1930s, were eager to develop new applications for the new material. They did this by careful trials and experimentation with the use of little more than simple statics to help justify their designs. For the most part, the

Figure 15. Stand at Madrid race course (1935) by Eduardo Torroja.
[*Photo: Cement and Concrete Association.*]

drivers for these developments were economic rather than architectural and most concrete shells were used for industrial buildings, railway stations, airports and the like. The French engineer Eugene Freyssinet built the first of his many shell roofs at a glass factory in 1915 and by 1921 he had constructed the remarkable airship hangers at Orly, some 60 meters high and spanning over 85 meters. The corrugated reinforced concrete shells were just 90 millimeters thick. Sadly they did not survive American bombs in 1944.

Different concrete engineers developed or exploited shells in different ways and, unusually among engineering structures, individual styles can be recognized among the great exponents.

Luigi Nervi's domed roof over the large sports hall for the Rome Olympics in 1958 spans some 100 meters and the corrugated concrete shell is just 25 millimeters thick; it is fabricated from precast units made from Nervi's own "ferro-cement" in which the reinforcing "bar," or rather wire, is just a few millimeters in diameter (Figure 16).

Figure 16. Sports hall for Olympic Games, Rome (1958) by Luigi Nervi.
[*Photo: Cement and Concrete Association.*]

Figure 17. Restaurant Los Manantiales at Xochimilco in Mexico (1958) by Felix Candela.
[*Photo: Cement and Concrete Association*.]

Figure 18. Market hall roof, Plymouth, by British Reinforced Concrete.
[*Photo: Bill Addis*.]

Felix Candela, on the other hand, was master of the hyperbolic paraboloid. This family of anticlastic curved surfaces have the great advantage that they can be generated using straight lines, making the timber formwork especially easy to construct. His remarkable restaurant Los Manantiales at Xochimilco in Mexico (1958) is made from his version of ferro-cement. It spans 42.5 meters and yet is just 42 millimeters thick, except at the free edges which are slightly thicker (Figure 17).

The majority of concrete shells built throughout the century were, of course, neither large nor architecturally memorable, and were often a standard product offered by contractors. However, many of these had an almost vernacular charm of their own, and were significant engineering feats. The roof over Plymouth market, built by British Reinforced Concrete in 1960, is one good example (Figure 18).

As shell sizes grew, it became necessary to understand better the stresses and deflections in shells. In the absence of useful engineering theory, engineers developed, from the 1930s to 1960s, increasingly sophisticated techniques for testing small models of shell roofs and scaling up the predictions a hundred or more times for the real thing (Figure 19). As in other areas of structural engineering, techniques of model analysis have largely (but not entirely) been replaced by computer-based methods, and the complex shell roof of the Sydney Opera House (1957–1962) was probably the first example of a building structure that could not have been constructed without the analytical work done using what were then the largest computers in the world (with the power of today's pocket calculator). Dozens of alternatives were analyzed before choosing one that satisfactorily met all the various constraints.

Just as computing power was unleashed in the 1960s, however, shell roofs began their demise for commercial reasons—their demand for enormous quantities of temporary supporting structures made them time consuming and labor intensive. Generally speaking they lost out to long-span steel structures and, from the 1970s, a variety of tension structures using cables and taut membranes. Nevertheless, shells still have their place when the circumstances are right. The Swiss engineer Heinz Isler, for instance, has enabled his shells to remain competitive since the 1970s in the face of strong competition from steel alternatives by developing a particularly ingenious system of reusable formwork (Figures 20 and 21).

See also **Buildings, Prefabricated; Concrete, Reinforced; Dams**

Bill Addis

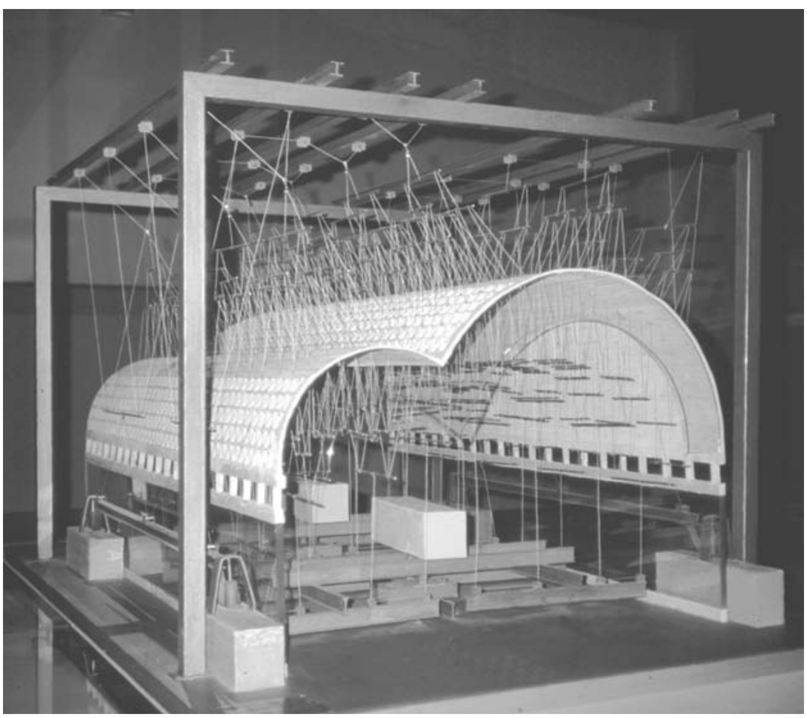
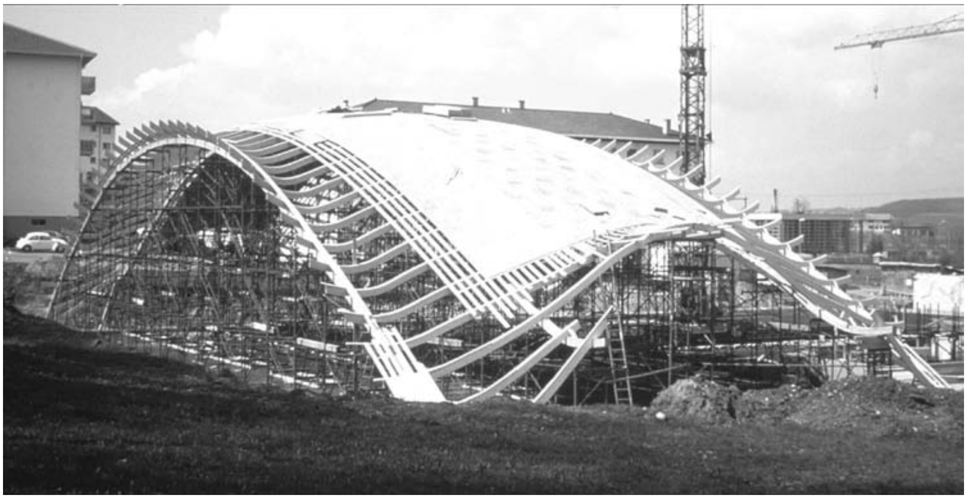


Figure 19. Model used in design development of the roof over the 60-meter-long pelota court at Recoletos, Spain, by Eduardo Torroja. The 1:200 scale model helped confirm the structural behavior of the 80-millimeter-thick shell spanning some 32.6 meters.
[*Photo: Bill Addis.*]



Figures 20 & 21. Reusable formwork for a shell roof over an indoor tennis court, near Norwich, England by Heinz Isler.
[*Photos: Heinz Isler, Tony Copeland.*]

Further Reading

Anchor, R.D. Concrete shell roofs, 1945–1965. *Proc. Instn. Civ. Engrs. Structs. Bldgs.* 116 381–389, 1996.

Chilton, J. *Heinz Isler*. Thomas Telford, London, 2000.

Faber, C. *Candela: The Shell Builder*. Architectural Press, London, 1963.

Nervi, P.L. *Aesthetics and Technology in Building*. Harvard University Press, Harvard, 1966.

Mainstone, R.J. *Developments in Structural Form*, 2nd edn. Architectural Press, Oxford, 1998.

Morice, P.B. and Tottenham, H. The early development of reinforced concrete shells. *Proc. Instn. Civ. Engrs. Structs. Bldgs.*, 116, 373–380, 1996.

Smith, A.C. Concrete shells before 1930: technique and architecture. *International Association for Bridge and Structural Engineering (British Group)—Colloquium on History of Structures: Cambridge 1982*. Institution of Structural Engineers, London, 1982, pp. 137–142.

Torroja, E. *The Structures of Eduardo Torroja*. F.W. Dodge Corporation, New York, 1958.

Constructed World

The term "constructed world" has shallow and deep significations. In the shallow sense, it refers to a contingent assemblage of those artifacts that have in fact been fabricated by human beings. In the deep sense, the term suggests that the world itself as a unity may be taken to be a human construction. That the proportion of human

experience engaged with artifacts has, especially since the Industrial Revolution, been dramatically increasing, itself tends to promote a shift from the shallow to the deep meaning. Related terms include "built," "engineered," and "technological world" or "environment."

That what might be constructed is not just products, processes, or systems, but a whole world, is an idea of unique twentieth century provenance. Although its most prominent manifestations are undoubtedly in relation to technology, during the 1900s the concept of construction increasingly became the basis for interpretations of art, architecture, psychology, education, economics, politics, ethics, knowledge, and even mathematics. From the vantage point of such a comprehensive if eclectic constructivism, all of human history is prefatory to an ethos of world fabrication that has been influenced by and in turn influences contemporary technology.

Despite, or perhaps because of, the overwhelming prominence of human construction in the twentieth century—from consumer goods through buildings to cities, from macroscale projects such as the U.S. Interstate Highway system and the European Channel Tunnel to genetic engineering and nanoscale mechanics, also including the unintended anthropogenic impacts on global biodiversity and climate—there exists no systematic overview of the world as an artifact. Instead, the (intentional and unintentional) complexity of the constructed world has thus far been conceived only piecemeal through a plurality of analytic and reflective approaches, among them history, architecture, urban planning, product design, and a diversity of related issues.

History

The history of such humanoid constructions courses over a million-year trajectory in which artifice remained subordinate initially to direct relations with nature (in hunting and gathering cultures) and then to social organization (in the rise of those axial civilizations characterized by farming and literacy in Mesopotamia, Egypt, and India). This broad distinction between artifice subordinate into natural and social milieux remains defensible even when qualified by the evidence for large-scale human terraforming, perhaps unintentional, prior to the development of literacy.

Mythological assessments of human construction include the stories of Abel and Cain (Genesis 4), the Tower of Babel (Genesis 11), Prometheus, Icarus, and more. Philosophical efforts to assess the relationship began with Plato's critique of *techne* practiced independently of wisdom (*Gorgias*) and Aristotle's implicit distinction between cultivation and construction. For Aristotle, the primary *technai* are those that cultivate nature, thereby helping her bring forth more fruitfully that which she is in principle able to bring forth on her own: the arts of agriculture, medicine, and education. Of real but subordinate interest are the constructive arts that produce artifacts such as structures, roads, and ships. Indeed, one way to frame the trajectory of human history over the last 5000 years is from cultivation to construction.

Certainly modernity arose in the fifteenth century in part as a conscious attempt to privilege constructive invention over cultivation. Francis Bacon, among others, called not just for the cultivation of nature but its systematic transformation, and cited as paradigmatic inventions to be imitated the printing press, gunpowder, and the compass. Galileo Galilei and others likewise proposed an augmentation of the human senses by means of the telescope, microscope, and related scientific instruments. It is the new commitment to inventive reconstruction in both the laboratory and the world that formed the basis for an historical emergence two centuries later of the Industrial Revolution. Indeed, the twentieth century in particular has witnessed the instrumentalization of the human sensorium that began in the laboratory and went public to alter the means of communication in commerce, politics, and entertainment (telephone, motion pictures, radio, television, and the Internet).

This historically unprecedented degree of technical mediation by means of tools, machines, and information technologies undermines all efforts to apply to the twentieth century the characterization of previous epochs by reference to the distinctive material substrates (Stone Age, Bronze Age, Iron Age, etc.). Although proposals have been made to describe the 1900s as the age of electricity, the atomic age, or the computer age, in truth it is more accurate to define the century not in terms of some specific technology but simply as the technological age—with diverse and ever-diversifying technologies serving as multiple means of world construction.

Even more reflective of the distinctive twentieth century consciousness of the world as construction is the effort to complement retrospect with prospect to forecast what will happen next: technological change, if not progress. Futurology, with roots in prophetic sociology and science fiction,

has nevertheless proved largely ineffectual. Relying more on trend analysis and imagination, it fails to engage the constructors themselves or to bring under effective economic or political directives the operative means operative for shaping the future.

Architecture

Efforts to go beyond futurology to develop a systematic analysis of the constructive elements in human affairs that might engage political and economic power grew out of the tradition of reflective building that finds classic expression in *De Architectura* by Vitruvius (circa 90–20 BC). Originally architecture designated the art of the master builder of the primary structures of the city (temples, palaces, monuments) and the layout of urban spaces in a manner that would reflect cultural ideals about the cosmic place and relations of humans. According to architectural historian Vincent Scully (1991) human builders have two basic options: to imitate natural forms or to oppose them. Compare, for instance, the architecture of indigenous peoples of the southwestern U.S., whose horizontal and earth-toned pueblos blend into a landscape defined by geological sedimentation and erosion, with the vertical thrust of those archetypical twentieth century buildings known as skyscrapers that dominate the cityscapes of Chicago or New York. On the ground, likewise, the modern city is laid out not to conform with a typology and the variegated paths of animal ambulance but as a block grid that extends into an instrumentally surveyed countryside, imposing simplification and legibility over the complex and intimate contours of rivers and mountains. Indeed, as Mumford (1961) states, as the constructed world became more and more extensive, the "city that was, symbolically, a world" was superceded by "a world that has become, in many practical aspects, a city."

Twentieth century transformations in the architecture of the constructed world have been driven by changes in materials, energy, transport and communication, and the commodities of peace and war. The first three achieved during the mid-1900s the apotheosis of developments with roots in the Industrial Revolution. Traditional construction materials such as wood and brick first became standardized and mass produced (e.g., dimensioned lumber), and then superceded as structural elements by iron, steel, and reinforced concrete; coal as an industrial energy source was complemented by oil, gas, and then nuclear power, with energy distribution and end-use itself accumulating from the mechanical and chemical to the electrical and electronic; alongside pre-twentieth century boats and railroads there moved with increasing speed and numbers the inventions of automobiles and airplanes, while communication networks competed with those of transportation to make human world construction a dynamic planet-covering web. The 1960s images of the earth from space, with lighted continents and pollution plumes, visually defined the paradox of multiple-scale human dominance and its responsibilities—even, some argued, its limits.

Focusing first on the static aspects of this dominance, structural engineer David Billington (1983) has analyzed the influence of the new materials of steel and reinforced concrete on structures. For Billington, twentieth century structures are defined by the intersection of three factors: efficiency, (i.e., the scientifically guided pursuit of minimal materials use); economy, the market-monitored effort to reduce monetary cost; and the understated achievement of elegance through maximum symbolic expression (given the least amount of materials and cost). In structures of spare democratic utility such as bridges, tall buildings, and free-spanning roofs over industrial workplaces and warehouses, aircraft hangers, and sports complexes, architectural engineers came into their own.

> "Structural designers give form to objects that are of relatively large scale and of single use, and ... see forms as the means of controlling the forces of nature to be resisted. Architectural designers ... give form to objects that are of relatively small scale and of complex human use, and ... see forms as the means of controlling the spaces to be used by people" [*D. Billington, 1983, p. 14*].

Bridges can be designed by engineers without architects; houses by architects without engineers. The engineered integration of efficiency and economy is realized in an esthetic of structural simplicity and thinness, as illustrated by the prestressed concrete bridges of Robert Maillart in Switzerland, the exposed steel tube x-bracing of Fazlur Kahn's John Hancock Center in Chicago, and the ribbed-concrete dome of the Palazzetto dello Sport by Pier Luigi Nervi in Rome.

Unlike structural engineering, early twentieth century architecture was less able to achieve an esthetic integration of science and democratic commerce, in part because it had to contend with well-established traditions of symbolic expression of the built world: the political iconography of Greek and Roman columns, the religious expres-

sion of the church spire, the solid façade of the bank, the decoration of Victorian domesticity. As the world-city emerged, architecture found itself caught in a cross-fire between scientific rationalism, industrial commercialism, and poetic romanticism—unclear which way to turn. The fundamental choice appeared to be between acceptance of technology or opposition to it. The winning synthesis was to take the scientifically rationalized artifact, that is, the machine, as an ideal for commercial exploitation and esthetic adaptation. In the architectural profession—itself now internally split into engineer, architect, and construction worker—this synthesis became a search for ways to design buildings that organized space in such a way as to parse human interactions into appropriate routines and to reduce resistance to their rapid interactions while minimizing the labor of construction of buildings for assembly lines, business offices, and large urban populations. The uniquely twentieth century architecture of these ubiquitous constructions, so named by a 1932 exhibition at the New York City Museum of Modern Art, was an "International Style" whose principles were an emphasis on "volume rather than mass," "regularity rather than axial symmetry," and the proscription of all "arbitrary applied decoration." This style, also known as modernism, was the first truly original building form since the rise of twelfth century Gothic.

The international style rejects the building patterns of premodern cultures (Greek, Roman, Gothic) in favor of shapes grounded in the efficient use of new materials and energies. Although steel and concrete were used initially to imitate Roman columns and Gothic arches, just as electric lights were first made to look like candles or gas lamps, in short order both became a flexible means for the design of indeterminate space and openness instead of determinate mass and enclosure. Geometric simplicity stripped of all ornamentation and standardized in modular forms at all levels, from structural members to external façade and finishing elements, contributed both to ease of construction and functional utilization.

Two leaders of this international modernism were Walter Gropius and Le Corbusier. Gropius, as the director of the Bauhaus in Germany, an engineering and product design school of great influence, eagerly embraced the machine esthetic in both buildings and their furnishings. Le Corbusier likewise condemned traditional building, redefined the house as "a machine for living in," and promoted the construction of whole cities of high-rise concrete apartment houses in repeating blocks connected by open roadways. The high-rise building made possible by the steel frame and electric elevator became a progressively simplified form, as illustrated by the now destroyed World Trade Center towers in New York and the Sears Tower in Chicago, emblematic of that modernist international architecture that dominated the first half of the twentieth century.

Without wholly rejecting the international style, the second half of the century nevertheless witnessed a rising attraction of more complex and interesting architectural spaces—an attraction most visually manifest in a postmodern ironic complexity that playfully revived traditional forms layered over the retained modernist structural elements. The popularity of postmodernism had, however, a counterpoint in the discovery and defense of vernacular architecture.

Urban Planning

As indicated, the constructed world consists not just of structures designed by architects but of cities, including urban and suburban systems, linked with transportation and communication networks across landscapes constructed for farming, recreation, and preservation. Although architecture classically included issues of city design, urban planning has in the twentieth century become an independent profession, due to the manner of engineering and construction work.

At the beginning of the century, urban planner Ebenezer Howard proposed a vision of the garden city at odds with what would emerge as the international style. For Howard the problem of increased urban population was not to be solved simply by efficient modular housing inspired by the standardization and interchangeability of parts and machine construction, but by recognizing what he called the "twin magnets" of the town and the country. The benefits of towns are high wages, sociability, and culture, yet at the cost of high prices and congestion. The countryside is the source of natural beauty and quiet, at the risk of boredom and lack of aspirations.

> "But neither the Town magnet nor the Country magnet represents the full plan and purpose of nature . . . . Town and country *must be married*, and out of this joyous union will spring a new hope, a new life, and new civilization" [*E. Howard 1965 [1902], p. 48*].

This utopian vision became a major basis for criticism of the rationalist esthetic of high modernist architecture. Whole new small, mixed-use towns exhibiting an interweave of superblocks with nar-

rower loop streets and cul-de-sacs instead of the repeating box grid were actually constructed in, for instance, Letchworth and Welwyn, England, and Radburn, New Jersey. Such experiments failed to live up to their promises of creating truly self-sustaining communities, as they became enclosed by larger suburban sprawl. Other influences of the garden city ideal can nevertheless be found in landscape architecture and the design of major urban parks, not to mention the construction of state and national parks and forests in both the U.S. and Europe, and eventually throughout the world.

The most practical innovation of early twentieth century urban planning was, however, the establishment of zoning laws that allowed for the political regulation of building practices. By the middle of the century architects and city planners were increasingly working together, with efforts also being made to enhance democratic participation in urban planning. The more grandiose schemes of Le Corbusier (who proposed a rebuilding of Paris) or Robert Moses (the New York state and city official who controlled its park and transportation development for more than 30 years), were moderated by local interests. Between them, social critics such as Jane Jacobs (1961) and urban planners such as Constantive Doxiadis (1963) brought realism and a more inclusive or interdisciplinary holism to thinking about the constructed world on the larger scale. The last half of the century also witnessed a new awakening of efforts among planners to take the natural environment into account in urban planning. Here the work of Ian McHarg (1969) exercised formative influence.

Product Design

Parallel to the architectural development of a machine esthetic at the level of structures, in tension with the organic ideals of urban planners, the commodities of peace and war were undergoing their own constructive transformations. Tools (dependent on human energy and guidance) were increasingly complemented if not replaced by machines (driven by nonhuman energy but still directed by human agents) and eventually semi-autonomous machines (requiring only indirect human guidance via feedback systems or programs), with the tools to machines transition continuing from the nineteenth century and dominating during the first half of the twentieth, and the rise of automation highlighting the second half. Distinctive of the century as a whole was the construction of a new type of household commodity—the electrical appliance—and then the electronic tool-machine represented most popularly by radios, televisions, and computers.

Prior to the rise of modern technology, the design of artifacts serving daily life was embedded in the craft of making—a virtually universal activity. Almost everyone was an artisan in the home, workshop, or field, and thus at one and the same time a person who conceived, fabricated, and used the indigenous basics of material culture. People "designed" things in the course of constructing them, so that making seldom involved any substantial moment of thinking through or planning beforehand, but proceeded as intuitive cut-and-try fabrication, guided by indigenous materials, traditions, and community. What has come to be called consumer product testing took place right in the making and immediate using by the maker, with the result that the commodities from regimes of craft production typically exhibit a certain practical artistic quality and honesty.

The Industrial Revolution's replacement of human power with coal- and steam-driven prime movers, its gearing of power into repetitive motion, and the required divisions of labor in manufacture, brought forth two needs: (1) the need for the designer as standard pattern maker so that artifacts could be mass produced; and (2) a need for the designer as style giver so that they could be mass marketed. Such a separation of design from construction and use could not help but open the door to a qualitative decline in the commodities produced, in reaction to which there emerged diverse efforts to reintroduce "art" into the new regime of industrial production; that is, to reunite what had been separated.

In the early stages, various arts and crafts movements sought to revive aspects of preindustrial modes of production, but at the beginning of the twentieth century the industrial design movement took a different approach, applying to quotidian commodities the principles being pursued in modernist architecture. Indeed, Gropius at the Bauhaus promoted modernist, technological simplification both in buildings and in streamlined furniture (see the famous Marcel Breuer chair). As one leading historian of product design has summarized the movement:

> "By the end of the Second World War, the practice of styling mechanical and electrical goods to make them appear clean, crisp, geometrical and, above all, modern, had become commonplace. Cars, electric razors, radios, food-mixers, typewriters, cameras, washing-machines, and so on, were all given body-shells reflecting the

machine esthetic of efficiency and functionalism" [*P. Sparke 1986, pp. 49–50*].

In the last half of the century, however, in product design as in architecture, questions arose about notions of rational objectivity and universality, especially in a market dependent on advertising. The psychological requirements of the mass consumer were granted increasing legitimacy, so that expendability and playful symbolism began to replace stricter rationalisms. In counterpoint to a culture of waste and simulacra however, designers such as Victor Papanek called first for a new applied realism (1971) and then respect for the ecological imperative (1995) in product design. The question of sustainability emerged in relation to both human markets and the natural environment.

In summary, the constructed world is a historical phenomenon that has during the twentieth century emerged on three levels: the intermediate level of buildings or structures (architecture), the large-scale level of cities and landscapes (urban planning), and the small-scale level of consumer goods (product design). There are nevertheless other levels of and perspectives on construction that have been passed over here: the microlevel construction in biotechnology and genetic engineering and nanoscale engineering design, politics and warfare (construction through destruction), the economics of globalization, information technology and the construction of the networked world, and the multiple media-based transformation of life and leisure. There also remains the need for a broadly based, general understanding of construction that would unite such levels and approaches.

See also **Bridges; Construction Equipment; Concrete Shells; Dams; Highways; Skyscrapers; Tunnels and Tunnelling**

CARL MITCHAM

Further Reading

Alexander, C. *The Timeless Way of Building*. Oxford University Press, New York, 1979. Attempts to think the unconscious principles of vernacular design.

Billington, D.P. *The Tower and the Bridge: The New Art of Structural Engineering*. Basic Books, New York, 1983.

Coates, J.F., Mahaffie, J.B. and Hines, A. *2025: Scenarios of US and Global Society Shaped by Science and Technology*. Oakhill Press, Greensboro, NC, 1997.

Doxiadis, C.A. *Architecture in Transition*. Oxford University Press, New York, 1963.

Gropius, W. *The New Architecture and the Bauhaus*. MIT Press, Cambridge, MA, 1965. Original German edition 1935.

Hitchock, H.-R. and Johnson,P. *The International Style*. W.W. Norton, New York, 1966. First published as *The International Style: Architecture Since 1922*. Museum of Modern Art, New York, 1932.

Howard, E. *Garden Cities of To-Morrow*. MIT Press, Cambridge, MA, 1965. First published 1902. Proposes satellite cities separated by green belts as a solution to dehumanizing mass urbanization.

Jacobs, J. *The Death and Life of Great American Cities*. Random House, New York, 1961. Argues that modernist architectural rationalism destroys natural urban growth and meaning.

Jencks, C. and Kropf, K., Eds. *Theories and Manifestoes of Contemporary Architecture*. Academy, Chichester, 1997. A collection of texts reflecting the explosion of architectural theory that characterized the last half of the twentieth century.

Le Corbusier [pseudonym of Jeanneret-Gris, C.E.] *Towards a New Architecture*. Brewer & Warren, New York, 1927. Original French edition 1923). The classic statement of the machine esthetic of modernist architecture.

Margolis, J. *Historied Thought, Constructed World: A Conceptual Primer for the Turn of the Millennium*. University of California Press, Berkeley, CA, 1995.

Mumford, L. *The City in History: Its Origins, its Transformations, and its Prospects*. Harcourt Brace Jovanovich, New York, 1961.

Papanek, V. *Design for the Real World: Human Ecology and Social Change*, 2nd edn. Van Nostrand Reinhold, New York, 1984.

Papanek, V. *The Green Imperative: Natural Design for the Real World*. Thames & Hudson, New York, 1995.

Rudofsky, B. *Architecture without Architects: A Short Introduction to Non-Pedigreed Architecture*. Doubleday, Garden City, NY, 1964. The catalog of a photography exhibit at the Museum of Modern Art in the mid-1960s.

Scott, J.C. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven, CT, 1998. Describes how modernist construction has manifested itself in politics and social planning.

Scully, V. *Architecture: The Natural and the Manmade*. St. Martin's Press, New York, 1991. An interpretive overview of Western architecture.

Sparke, P. *An Introduction to Design and Culture in the Twentieth Century*. Allen & Unwin, London, 1986.

Thomas Jr., W.L. Sauer, C.O., Bates, M. and Mumford, L. Eds. *Man's Role in Changing the Face of the Earth*, 2 vols. University of Chicago Press, Chicago, 1956. Collects over 50 papers by major scholars from an extended collaboration. Vol. 1 provides historical background; vol. 2 analyzes the processes by which humans are modifying the water, climate, soils, biota, as well as the ecology of wastes and urbanization. Includes extensive discussion summaries.

Venturi, R., Brown, D.S. and Izenour, S. *Learning from Las Vegas*. MIT Press, Cambridge, MA, 1972. Architecture can move beyond modernist formalism by appreciating the human pleasures of a playful commercialism.

Construction Equipment

The twentieth century had barely begun when 102 American-made steam shovels began excavating the Panama Canal. Those coal-fired giants devised in the nineteenth century for building railroads

were one of the few large-scale earthmoving tools available in 1904. They were massive in size and weight and required a team of workers to operate. While heavy chains transmitted power to the excavator's bucket, the rail-mounted unit itself was not self-propelled, but had to be towed from work site to work site. Subsequent earthmoving machines evolved quite differently.

Although the focus of much twentieth century construction work was on road building, there was foundation work for buildings of all sizes, as well as civil engineering projects such as dams. The horse-drawn graders and scrapers used for leveling work on these undertakings during the first decades of the century had changed little from their nineteenth century origins. The first entirely new machine to appear was a tractor, which moved on crawler tracks and was used for towing earthmovers. It evolved from a wheeled, gasoline-powered agricultural tractor designed by Benjamin Holt in 1908 for use on the soft farmland of California.

During the 1920s the versatility of the crawler tractor was enhanced by the addition of a front-mounted scraper blade. Thus, the tractor itself became an earthmover. The technique of moving earth by a blade pushed from behind came directly from an earlier horse-powered device known as a bulldozer. It was from this that the name for the tractor and blade combination was derived. Blades were still accessories in the 1930s and their manufacture was, for the most part, by independent companies that supplied the tractor builders. The first hydraulically controlled blade appeared in 1925. But despite the compactness and precision offered by that system, the industry was slow to adopt the improvement and cable control remained the norm until the late 1930s. During the late 1920s, the efficiency of blades was improved by the addition of teeth to their leading edge. It was almost another decade before they were adopted industry wide. The decade closed with the introduction of the self-propelled grader.

The 1930s was an active decade in the earthmoving industry and there were several watershed events. None had greater impact than the introduction of diesel power. Although high-speed diesel engines were mass-produced in Germany during the late 1920s for tractor and lorry propulsion, an American-made diesel engine changed the industry. It was the first mass-produced diesel engine developed, manufactured, and applied as power in a mass-produced vehicle—the Caterpillar tractor. Diesel engines were superior to gasoline and oil power plants for use in heavy machinery for several reasons. Not least of all was the concentrated weight of the diesel engine, which was an asset in earthmovers where traction was a factor of weight. The engines provided power and lugging capacity not found in any other power source.

Despite their practicality, bulldozer tractors were most efficient when operated over relatively short distances and they were capable of moving only relatively small volumes. When a number of cubic meters of earth were to be moved over a distance of more than a few meters, it was best done by a scraper or bottom-dump trailers or rear-dump trucks. The other significant event of that era occurred when the metal tires on these vehicles were replaced with pneumatic heavy-duty rubber tires. Not only was the equipment more manageable on rough terrain, but its potential top speed increased. Nowhere was this more significant than on scrapers, which in 1938 were revolutionized with the introduction of the diesel-powered self-propelled LeTourneau Tournapull wheel tractor and scraper. This followed a decade-long trend in earthmoving machinery toward increased capacity and horsepower.

During World War II earthmoving machinery proved especially useful in the speedy preparation of supply depots, aircraft landing strips, and fortifications. The end of the war marked a turning point in the industry. In the immediate aftermath of the war, recovering Europe provided a ready market for American machinery exports. To meet the growing demand, some American companies invested in plants there. As conditions improved during the 1950s, a vital and creative domestic earthmoving equipment-manufacturing industry developed. Companies such JCB and Priestman in Britain and Atlas and Demag in Germany were among the many that arose throughout Europe.

The hydraulically operated excavator—a descendant of the steam shovel and the succeeding power shovel—was introduced in Germany in 1954. Up to that time, the control functions of power shovels were through cables. The industry's embrace of the excavator with components roughly analogous to the human arm and hand and a fluidity of movement to match was so thorough that power shovels were no longer used as a construction tool.

Of the many versatile machines developed during the early 1950s, the wheeled loader—also known as a front-end loader, bucket loader, or tractor shovel—was an immediate and widespread success. The nimble and highly maneuverable rubber-tired tractor with front-mounted hydraulically controlled bucket could be used to dig, lift, and quickly fill waiting dump trucks. The versati-

lity and value of these machines increased tremendously in the mid-1950s when JCB in Britain and Case in the U.S. marketed factory-made units in which tractor loaders were joined with the boom, dipperstick, and bucket of the backhoe. The loader or backhoe became the most widely used tool on small-scale building projects.

American companies, which dominated the industry until the 1960s, faced a steady increase in competition from both European and Japanese machinery builders during that decade. By the late 1960s, machines were reaching the practical limit in their size and during the 1970s builders devoted their efforts toward improved equipment productivity. This was accomplished through increased horsepower, refinements in hydraulic controls, easier serviceability, and operator comfort.

Developments in earthmoving technology were not always embraced simultaneously or universally. Hydrostatic drive, in which fluids transmitted power to the wheels or tracks, was first used in Europe during the 1960s and by the mid-1980s was almost standard on all smaller machines built there. Large American manufacturers preferred mechanical drive, which had long proven its reliability and cost effectiveness. During the 1980s there was a general adoption of high-drive sprockets for crawler track tractors. Through modifications to the drive system and the addition of a third sprocket, the drive for crawlers was moved from ground level where it was subject to damage, to a safer and more easily serviced elevated position. The operation of hydraulic systems was optimized by the incorporation of electronics. Finally, in the 1990s a variety of diminutive-sized machines such as the mini-excavator were developed which all but eliminated the need for handwork at some job sites.

See also **Power Tools and Hand-Held Tools**

WILLIAM E. WORTHINGTON, JR.

Further Reading

Grimshaw, P.N. *Excavators*. Blandford, Poole, 1985.

Grosser, M. *Diesel: The Man and the Engine*. Atheneum, New York, 1978.

Harris, F. *Construction Plant: Excavating and Materials Handling, Equipment and Methods*. Garland, New York, 1981.

Haycraft, W.R. *Yellow Steel: The Story of the Earthmoving Equipment Industry*. University of Illinois, Urbana, 2000.

Leffingwell, R. *Caterpillar*. Motorbooks International, Osceola, 1994.

Oberlender, G.D., Ed. *Earthmoving and Heavy Equipment*. American Society of Civil Engineers, New York, 1986.

Young, J.A. and Budy, J. *Endless Tracks in the Woods*. Crestline, Sarasota, 1980.

Contraception, Hormonal Methods and Surgery

The discovery of hormones in 1898 and subsequent developments in biological science during the twentieth century were necessary before hormonal methods of contraception could be made available. Physical barrier methods of birth control and douching after intercourse had been tried with varying degrees of success for centuries, but the idea of preventing pregnancy with a pill was a dream of such women's health advocates as Margaret Sanger, founder of the Planned Parenthood clinic

The history of the development of oral contraceptives must include the work of scientists in Japan in 1924 and Austria in 1927 who devised what became known as the "rhythm method" of birth control. In independent research, both groups realized that a woman's fertile period is approximately midway in the menstrual cycle (that is, counting from the first day of her period until the beginning of the next period) and that pregnancy could be avoided by abstaining from sex at that time. The connection between hormone levels and ovulation was clearer when scientists at the University of Rochester in New York identified the ovarian hormone progesterone in 1928 and recognized its importance in preparing the uterus for the implantation and sustaining of pregnancy.

The next steps included the isolation of estrogen by Edward Doisy of Washington University, St. Louis, Missouri in the 1930s and the discovery of a way to make synthetic progesterone by chemical professor Russell Marker in 1941. This would become the basis for hormonal birth control. Following the Food and Drug Administration (FDA) approval of Enovid in 1957 as a treatment for severe menstrual disorders, it was only three years until the manufacturer Searle received FDA approval to market Enovid as an oral contraceptive, and it was quickly named "the pill." Within five years over 6.5 million women were taking it, and oral contraceptives had become the most popular form of birth control in the U.S., resulting in a revolution in birth control technology that expanded women's reproductive choices.

There are a number of hormonal contraceptives, including the combined pill; the progestin-only pill (POP), or minipill; hormonal injections such as Depo-Provera; contraceptive implants such as

Norplant; patches such as Ortho-Evra; or the Nuvaring, a combination of the hormones estrogen and progestin. Each can be highly effective, if used according to instructions. Similarly, surgical methods such as tubal ligation for women or vasectomy for men can be equally effective, and both procedures became more widespread in the last three decades of the twentieth century.

Oral contraceptives (OCs) refer to pills containing both estrogen and progestin, although there are "minipills," which contain only progestin. When taken consistently, pills can prevent ovulation, thereby eliminating the midcycle pain that some women experience at the time of ovulation. Pills can also decrease menstrual bleeding, thus decreasing the likelihood of iron deficiency anemia. Birth control pills come in packs of either 21 "active" pills, or 28 pills, with 21 "active" pills and 7 placebo pills, designed to keep the user in the habit of taking a daily pill even during her period. It is also possible to prevent a cycle by taking extra pills from a separate package. Some forms of the pill have been proven to improve acne conditions.

Oral contraceptives have been determined to protect against some forms of cancer, including ovarian and endometrial cancer. In addition, the pill reduces anemia due to iron deficiency since it lessens the amount of blood lost during a woman's menstrual cycle. Despite these positive effects, the pill does not protect users from sexually transmitted diseases, and the pill's effectiveness rests on the user remembering to take to the pill every day.

Minipills, the progestin-only pills, are less effective than the combined OCs, since they completely change a woman's menstrual cycle, which can lead to a bloated feeling or increased weight gain. These pills contain a lower dose of progestin, and no estrogen; for these pills to be effective, users must take them at the same time every day. The progestin in these pills thickens the cervix mucus, making it difficult for sperm to enter the uterus or fallopian tubes. The risk for pregnancy is, however,greater for the minipill. For users of the combined pill, the risk of pregnancy is 3 percent, while for users of the minipill, the risk increases to 5 percent.

For those who have difficulty remembering to take the daily pill or for whom OCs are not recommended, there are other options, including Norplant, in which six 34-millimeter-long Silastic rods that release levonorgestrel are inserted into the woman's upper arm in a fanlike shape just under the surface of the skin. This method, although effective, can be expensive. The erratic bleeding patterns that result also may not appeal to many women; approximately 60 percent of Norplant users report irregular bleeding patterns within the first year of use.

Depo Provera injections, which are administered in the arm or buttocks every 90 days, are highly effective and fairly inexpensive, with an average cost of $40 per injection. Women receive Depo injections every 13 weeks, with each injection containing progestin, which thickens the cervix mucus. However, Depo shots can cause the same side effects as the pill and Norplant, often resulting in headaches, weight gain, nervousness, and dizziness. Depo shots have been proven to cause the most significant weight gain, with an average of 7.5 kilograms after six years of use. It is important that women receive these shots every 13 weeks for the shots to be effective; when used effectively, about 3 in 1000 will experience an unexpected pregnancy in the first year of use.

The most common surgical method of birth control for men is vasectomy. The procedure involves the removal of a portion of the vas deferens, thus resulting in male sterility; this procedure does not affect the male sex drive or ability to ejaculate. Less than 1 in 1000 couples will experience an unexpected pregnancy in the first year of sterilization; the sterility, however, is not immediate, and the male is still fertile for three months, or 20 ejaculations, after the completion of the vasectomy.

In the tubal ligation procedure for women, the fallopian tubes are tied off, and a section of each tube is removed, thus resulting in sterility. Eggs released from the ovaries each month are blocked from reaching the uterus, thereby preventing fertilization by sperm. Less than 1 in 100 couples will experience an unwanted pregnancy in the year following a tubal ligation.

See also **Contraception, Physical and Chemical Methods**

JENNIFER HARRISON

Further Reading

Davis, A.J. Advances in contraception. *Curr. Reproduct. Endocrinol.* 27, 3 597–610, 2000.

Hatcher, R. *et al. Contraceptive Technology*, 16th revised edn. Irvington Publishers, New York, 1994.

Useful Websites

Health and Wellness Resource Center. Contraception: http://galenet.galegroup.com/

Statement on Hormonal Methods of Contraception. International Planned Parenthood Federation. IMAP

Statement on Barrier Methods of Contraception: http://www.ippf.org

For history of the development of contraceptive methods, especially hormonal methods, see PBS online, American Experience: http://www.pbs.org/wgbh/amex/pill/timeline

Contraception, Physical and Chemical Methods

Contraception refers to the deliberate means of preventing pregnancy by interfering with the normal processes of ovulation, fertilization, and implantation of the sperm. The process that could eventually lead to pregnancy begins with the maturation of the ovum (egg), thereby preparing the lining of the uterus for the fertilized egg. Birth control methods can prevent this process; however, no type of birth control is 100 percent effective. Failure rates for contraceptive methods can be divided into "perfect use" and "typical use," the former derived from clinical trial data of highly motivated patients who have received support and reminders from research personnel, and the latter from reports from the average user.

The use of birth control methods has a varied history. The 1873 Comstock laws in the U.S. prevented the distribution of sexually explicit materials, which would include the purchase of condoms or spermicides from Europe. However, it was not until the efforts of Margaret Sanger and her sister Ethel Byrne in opening a clinic to dispense contraceptive devices in 1916 that the revolution in birth control methods began. Although the New York City Vice Squad shut down the clinic, Sanger's efforts resulted in a victory for birth control by giving physicians the right to provide contraceptive measures to their patients. The so-called sexual revolution of the 1960s and the women's rights movements, particularly in the U.S. and the developed nations gave further impetus to the attempts by women to better understand and control pregnancy and family size. Controversy surrounding these issues continued into the twenty-first century.

Physical and chemical methods of birth control can be divided into three categories:

1. Physical barrier methods, including the condom, diaphragm, and cervical cap, which work to prevent the sperm from getting to and fertilizing the egg (only the condom protects against sexually transmitted diseases).
2. Chemical barrier methods, such as spermicides, which kill sperm on contact; most contain nonoxynl-9, are placed in the vagina, work best in combination with a barrier method such as a condom, and come in the form of jelly, foam, tablets, or transparent film.
3. Intrauterine devices (IUDs), which are inserted into the uterus and can remain for up to ten years; they work to prevent a fertilized egg from implanting in the lining of the uterus but may also have negative effects.

Significant developments in effective physical methods of contraception in the nineteenth century date back to 1832 when Massachusetts physician Charles Knowlton promulgated douching after intercourse using a syringe with water-based solutions to which salt, vinegar, chloride, or other elements would be added. In 1838 the German physician Friedrich Wilde offered his patients the "Wilde cap," a small cervical cap that would be placed over the cervix between menstrual periods. Although of limited success at the time, his device is considered the precursor of the diaphragm. Charles Goodyear, who invented the technique to vulcanize rubber in 1839, also provided the means to manufacture rubber condoms, douching syringes, and diaphragms.

Perhaps the most common contraceptive methods in the twentieth century are the physical and chemical barrier methods, and when used in combination, these methods are highly effective. Physical barrier methods, such as condoms, diaphragms, and cervical caps, are the only methods that protect against sexually transmitted infections, although the protection is the greatest with condoms. Chemical methods of contraception, such as spermicides, are available in foam, cream, jelly, film, suppository, or tablet form. Inserted into the vagina, spermicides contain a chemical that destroys sperm. Some types of spermicide require a ten-minute waiting period before intercourse, and must remain in the vagina for at least six hours after intercourse.

Condoms, which come in both male and female forms, work to keep the sperm and egg apart. If used correctly, male condoms only result in a pregnancy rate between 3 and 14 per 100 women per year. The male condom, a sheath that covers the penis during sex, is usually made of latex, but for those with latex allergies, condoms made with synthetic materials such as polyurethane are also available. However, only latex condoms have been proven to be highly effective in preventing sexually transmitted infections. Some condoms contain spermicide, which may provide additional contraceptive protection. The female condom, a polyur-

ethane pouch with two flexible rings, one of which is inserted into the vagina, while the outer ring rests on the labia during intercourse, is less effective than the male condom. The Reality female condom was approved for use in the U.S. in 1993, and although it may provide some protection against sexually transmitted diseases, it is nowhere near as effective as the male latex condom. The estimated yearly failure rate for the female condom ranges from 21 to 26 percent, compared to a failure of 15 percent failure rate for the male condom.

Diaphragms, which must be fitted by physicians and women's health providers, should not be inserted longer than six hours before coitus and must be left in the vagina for at least six hours after coitus, but no longer than 24 hours. The diaphragm, a flexible rubber disk with a rigid rim, is designed to cover the cervix during and after intercourse so that sperm cannot reach the uterus. Spermicidal jelly or cream must be inserted inside the diaphragm for it to be effective. When used with spermicides, the diaphragm has a failure rate of 6 to 18 percent. There are some drawbacks, including the fact that diaphragms need to be kept clean and free of holes, and they should be refitted every other year. There is a chance that the diaphragm could become dislodged if the woman is on top during intercourse. Use of the diaphragm can also result in increased bladder infections.

Similarly, the cervical cap, which must be manually inserted by the woman, must also be used with a spermicide. Approved for contraceptive use in 1988, the cervical cap is a dome-shaped rubber cap that fits snugly over the cervix and must be fitted by a physician or women's health provider. Although it is more difficult to use, it may be left in place for up to 48 hours. There could be an increased occurrence of irregular Pap smear tests in the first six months of usage. The cap has a failure rate of 18 percent, and it may not be useful for all women since it is only available in four sizes and may be difficult to fit some women.

A less commonly used form of contraception, the intrauterine device (IUD), was first available in 1965 and has tended to result in the most satisfaction. Two forms of IUDs are available, a copper-containing IUD, the ParaGard T380a, marketed by Ortho-McNeil Pharmaceutical, Inc., and a progesterone-releasing IUD, Progestasert, produced by the Alza Corporation. The IUD interferes with sperm mobility and fertilization, although the timing of the IUD insertion is controversial. Printed literature states that an IUD should be placed in the vagina within five days of a menstrual cycle; however, an IUD can be inserted at any time a clinician is certain the patient is not pregnant. Ninety percent of IUD users are over the age of 35.

See also **Contraception, Hormonal Methods and Surgery; Fertility, Human**

JENNIFER HARRISON

Further Reading

Davis, A.J. Advances in contraception. *Curr. Reproduct. Endocrinol.*, 27, 3, 597–610, 2000.

Hatcher, R., *et al. Contraceptive Technology:* 16th revised edn. Irvington Publishers, New York, 1994.

Useful Websites

Health and Wellness Resource Center. Contraception: http://galenet.galegroup.com/

Statement on Barrier Methods of Contraception. International Planned Parenthood Federation, IMAP Statement on Barrier Methods of Contraception: http://www.ippf.org

For history of the development of contraceptive methods, see PBS online, American Experience: http://www.pbs.org/wgbh/amex/pill/timeline

Control Technology, Computer-Aided

The story of computer-aided control technology is inextricably entwined with the modern history of automation. Automation in the first half of the twentieth century involved (often analog) processes for continuous automatic measurement and control of hardware by hydraulic, mechanical, or electromechanical means. These processes facilitated the development and refinement of battlefield fire-control systems, feedback amplifiers for use in telephony, electrical grid simulators, numerically controlled milling machines, and dozens of other innovations.

Computational control in the decades before the 1950s usually meant cybernetic control involving automatic, closed-loop mechanisms. Massachusetts Institute of Technology (MIT) mathematician Norbert Wiener, who derived the word cybernetics from the Greek word for "steersman," considered the discipline—a direct forebear of the modern discipline of computer science—to be a unifying force created by binding together the theory of games, operations research, logic, and the study of information and automata. Wiener declared human and machine fundamentally interchangeable:

> "[C]ybernetics attempts to find the common elements in the functioning of automatic machines and of the human nervous system, and to develop a theory which will

cover the entire field of control and communication in machines and in living organisms." [*Wiener, 1948*].

Industrial, corporate, and government interest in fully electronic computer-aided control technology grew dramatically in the 1950s as access to electronic mainframe computers and transistor technology became more readily available. Indeed, information processing and computer programming emerged from the decade as virtual synonyms for control. In 1956, for instance, American Edmund Berkeley—editor of the journal *Computers and Automation* and founder of famed New York robotics company Berkeley Enterprises—defined control as the method by which one might "direct the sequence of execution of the instructions to a computer." Increasingly, computer control technology exploited open-loop systems, where a set of hard-wired commands or programmed instructions are executed in service to a predefined goal rather than uninterrupted governance.

American information technology management pioneer John Diebold explained in 1952 that realizing a truly automatic factory required "a machine that, once set up, performs a series of individual computations or steps in the solution of a problem, without further human intervention." Engineers touted the many advantages of computer hardware and increasingly *soft*-ware solutions in batch and continuous control operations, including the separation of function from material extension, mass fabrication, programmability, and flexibility. In the 1950s, computer controls found their way onto automobile assembly lines, railroad freight-sorting yards, and foundry grounds and into applications as various as packaging machines, furnace dampers, and iron-lung regulators. Early in this decade the Arma Corporation, an American Armed Forces contractor, introduced programmable machine tool automation—often referred to as numerical control or N/C—with its Arma-Matic lathe system. Later in the decade the Unimation ("Universal Automation") Company designed a programmable robot used to pull hot die-cast automobile parts out of their molds. In 1959 TRW and Texaco in Port Arthur, Texas, achieved one of the earliest installations of digital online process control in industry for the purpose of petrochemical catalytic polymerization.

In the 1960s computer control grew in importance in contexts beyond industrial production and logistics: office management, banking, air traffic control, passenger reservations, and biomedical diagnostics. The movement of computer control technology into previously sacrosanct workplace domains precipitated an outpouring of both positive and negative emotion. Among the ambiguous sets of social and cultural dilemmas weighed carefully during the control revolution of the late twentieth century were prospects for deskilling versus the pursuit of higher intellectual achievements; potential for mass unemployment versus information technology manpower shortages; increased managerial power versus decentralization of control; and inflexible working conditions versus profound achievements in workplace safety.

In the late 1960s and throughout the 1970s computer and cognitive scientists began experimenting with expert control systems, including neural networks, fuzzy logic, and autonomous and semiautonomous robots. Artificial neural networks incorporate an array of parallel-distributed processors and sets of algorithms and data operating simultaneously—presumably emulating human thought processes. The history of neural networks extends back to the cybernetic gaze of Wiener and the construction of the so-called McCulloch–Pitts neuron. This neural model, laced together with axons and dendrites for communication, represented a total cybernated system akin to the control processes of the human brain. Its namesake was derived from its American developers, the neurophysiologist Warren McCulloch and the mathematical prodigy Walter Pitts. Pitts in particular saw in the electrical discharges meted out by these devices a rough binary equivalency with the electrochemical neural impulses released in the brain. Thus it was but a short step from artificial to animal control and communication.

Fuzzy logic emerged at roughly the same time as neural network theory. The fuzzy concept—which allowed for intermediate degrees of truth rather than simple binary control—was first proposed by University of California, Berkeley computer scientist Lotfi Zadeh in 1965. Zadeh derived the idea from his study of imprecise human linguistics. Fuzzy logic is now usually subsumed under the artificial intelligence niche of subsymbolic artificial intelligence, and better known as "soft computing." Soft computing seeks control of haphazard, imprecise, and uncertain operations to form consensus opinions and make "semiunsupervized" decisions. Soft computing, a discipline still in its infancy, is currently focused on several areas of application, including computer-integrated manufacturing, computer-aided design, image processing, handwriting recognition, power system stabilization, and decision support.

In the 1980s a "long winter" of greatly reduced federal funding descended on artificial intelligence

research—including neural networks and fuzzy logic. Despite this, the 1980s and 1990s witnessed a general renaissance in robotic computer-aided control. The twentieth-century pursuit of robotic incarnations of the control process extends back to Karel Čapek's fictional play *R.U.R.* (Rossum's Universal Robots). In the 1950s, Berkeley began creating hundreds of "Robot Show-Stoppers," playful mechanical creations with electronic brains. Among them were Squee, a robot squirrel that gathered tennis ball "nuts," and James, an android department store greeter. Control of Berkeley's robots was achieved using mechanical relays, phototubes, contact switches, drive motors, and punched paper tape. Robots developed in the next decade included sophisticated tactile sensors and embedded heuristic programming simulating the rule-of-thumb reasoning of human beings.

Throughout the twentieth century enthusiasts and critics hotly debated the merits of intelligent robot control in human societies. By 1970 professional roboticists began turning away from robot control and toward robot assistance. Robots—including those acting at a distance (telerobotics)—remained firmly tethered to their human masters. Examples of semiautonomous robots of the last two decades of the twentieth century include the Martian rovers developed by the National Aeronautics and Space Administration's (NASA's) Jet Propulsion Laboratory and TOMCAT (teleoperator for operations, maintenance, and construction using advanced technology), which repairs high-voltage power lines without service interruptions.

See also **Artificial Intelligence; Computer Science; Software Engineering**

PHILIP L. FRANA

Further Reading

Anderson, J.A. and Rosenfeld, E. Eds. *Talking Nets: An Oral History of Neural Networks*. MIT Press, Cambridge, MA, 1998.

Berkeley, E.C. and Wainwright, L. *Computers: Their Operation and Applications*. Reinhold, New York, 1956.

The Computer Control Pioneers: A History of the Innovators and Their Work. Instrument Society of America, Research Triangle Park, NC, 1992.

De Wit, D. *The Shaping of Automation: A Historical Analysis of the Interaction Between Technology and Organization, 1950–1985*. Verloren, Hilversum, 1994.

Diebold, J. *Automation: The Advent of the Automatic Factory*. Van Nostrand, New York, 1952.

Goldberg, K., Ed. *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*. MIT Press, Cambridge, MA, 2000.

Grabbe, E.M., *et al.* Eds. *Handbook of Automation, Computation, and Control*, 3 vols. Ramo-Wooldridge, Los Angeles, CA, 1958.

Mindell, D.A. *Between Human and Machine: Feedback, Control, and Computing Before Cybernetics*. Johns Hopkins University Press, Baltimore, 2002.

Noble, D.F. *Forces of Production: A Social History of Industrial Automation*. Oxford University Press, Oxford, 1986.

Wiener, N. Cybernetics. *Scientific American*, 179, 14–19, 1948.

Zadeh, L.A. *Fuzzy Sets and Applications: Selected Papers*. Wiley, New York, 1987.

Control Technology, Electronic Signals

The advancement of electrical engineering in the twentieth century made a fundamental change in control technology. New electronic devices including vacuum tubes (valves) and transistors were used to replace electromechanical elements in conventional controllers and to develop new types of controllers. In these practices, engineers discovered basic principles of control theory that could be further applied to design electronic control systems.

The voltage and current regulation technology of electrical power networks was among the fields that witnessed an early transition from electromechanical to electronic control. In the early twentieth century, the most prevalent voltage–current regulator was the "Tirrill" regulator made by the General Electric Company. A Tirrill regulator had a constantly vibrating contact that moved a resistor in and out of a generator's excitation circuit. It used the presence or absence of the resistor to reduce or raise the output voltage. Electromechanical arrangements like the Tirrill regulator were gradually replaced in the 1930s by a fully electronic design using gas-filled thyratron tubes.

The thyratron was developed by General Electric in the 1920s. Known as the "grid-controlled rectifier," a thyratron had two modes of operation. When the grid voltage was less than a threshold, no current flowed between the anode and the cathode (off state). Once the grid voltage exceeded the threshold, the thyratron operated as a rectifier (on state). As long as the thyratron was triggered to the on state, it remained on until the voltage across the anode and cathode vanished. Thus unlike a vacuum tube whose conductivity continuously varied with the grid voltage, a thyratron kept the high-conductivity state after being triggered, no matter whether the grid voltage dropped below the threshold later. The average output voltage of a thyratron varied with the

duration of the on state, which was determined by timing of the rising edges of the voltage at the grid. A thyratron-based voltage regulator therefore adjusted the timing of the grid voltage according to the fluctuation of the output voltage (Figure 22).

A similar control device to thyratron was discovered in the age of semiconductors. Developed by M. Sparks, L.W. Hussey, and William Shockley at the Bell Telephone Laboratories in the early 1950s, a thyristor was a solid-state device constituted of a *p–n–p–n* structure. It had an anode connected to the first *p*-type block *p*1, a cathode to the second *n*-type block n2, and a gate to the second *p*-type block *p*2. Like the thyratron, the thyristor was a gate-controlled rectifier. It remained off when the gate current was inadequate to trigger its conduction. Once the gate current exceeded a threshold, the device suddenly transformed into a rectifier and remained in the on state. Therefore by adjusting the timing of the gate current's rising edges, a thyristor could also be used to construct a voltage regulator (Figure 23). With the same functional characteristics as thyratron tubes and the additional advantages of solid-state devices, thyristors eventually replaced thyratrons in power electronics.

The control technology of telephone networks was another early electronic system. A telephone network's typical control problem was how to make the output follow the time-variant input faithfully. The attenuation and dispersion of telephone signals were particularly serious when the communication distance was long. In the 1910s, the solution was to employ vacuum-tube amplifiers along a communication path to repeat signals, but these amplifiers had the problems of oscillation and nonlinear distortion. In 1923, Harold Black at the Bell Lab developed the concept of the negative feedback amplifier. In his design, the amplifier amplified the difference between the input and a feedback signal coupled from the amplifier's output (if the amplifier amplified the sum of the input and the feedback, then it would be a positive feedback amplifier). The negative feedback design reduced an amplifier's unstable resonance and extended the linear range of operation (Figure 24). Black was not the first person to come up with a negative feedback design, but he was the first to point out the fundamental difference between negative and positive feedback. Following Black, the Bell Lab engineers in the 1920s and 1930s conducted research on a mathematical theory of feedback amplifiers. Hendrik Bode and Harry Nyquist, for instance, developed representational tools, known as Bode plot and Nyquist diagram, for the design and analysis of feedback circuits.

The control in process industries (petroleum, chemicals, etc.) in the early twentieth century might not have been saliently affected by electronic technologies, but its improvement paved the way toward fundamental understanding of a control theory. An early process controller used either an electrical relay with a solenoid-operated valve or a motor-operated valve. In the first case, the electrical control signal turned the relay on or off, moving the valve in a two-stage fashion. In the second case, the control signal set the motor and

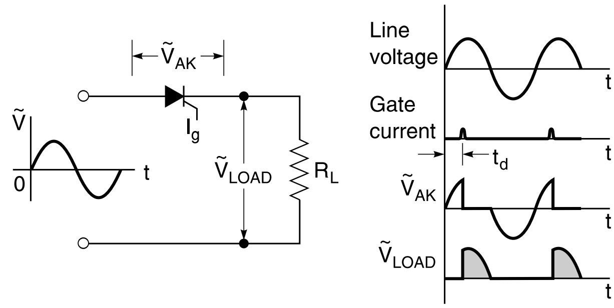


Figure 23. A thyristor regulator.
[*Source: Sze, S.M. Semiconductor Devices: Physics and Technology. Wiley, New York, 1985, p.153. © Wiley 1985. This material is used by permission of John Wiley & Sons, Inc.*]

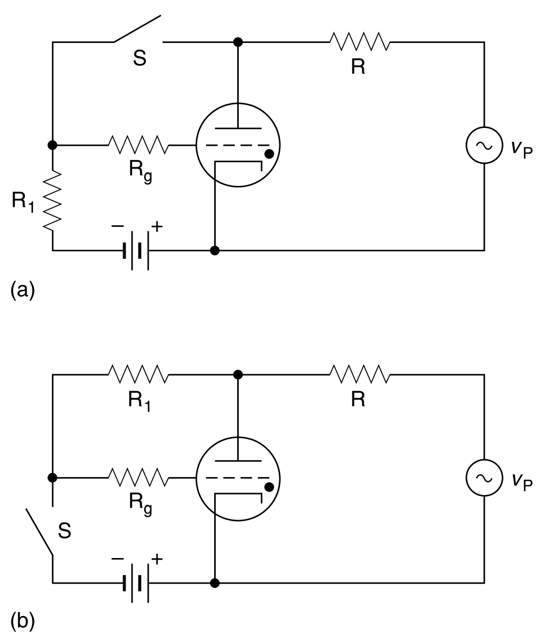


Figure 22. A thyratron control circuit.
[*Source: Happell, G.E. and Hesselberth, W.M. Engineering Electronics. McGraw-Hill, New York, 1953, p. 438. Reproduced with permission of The McGraw-Hill Companies.*]

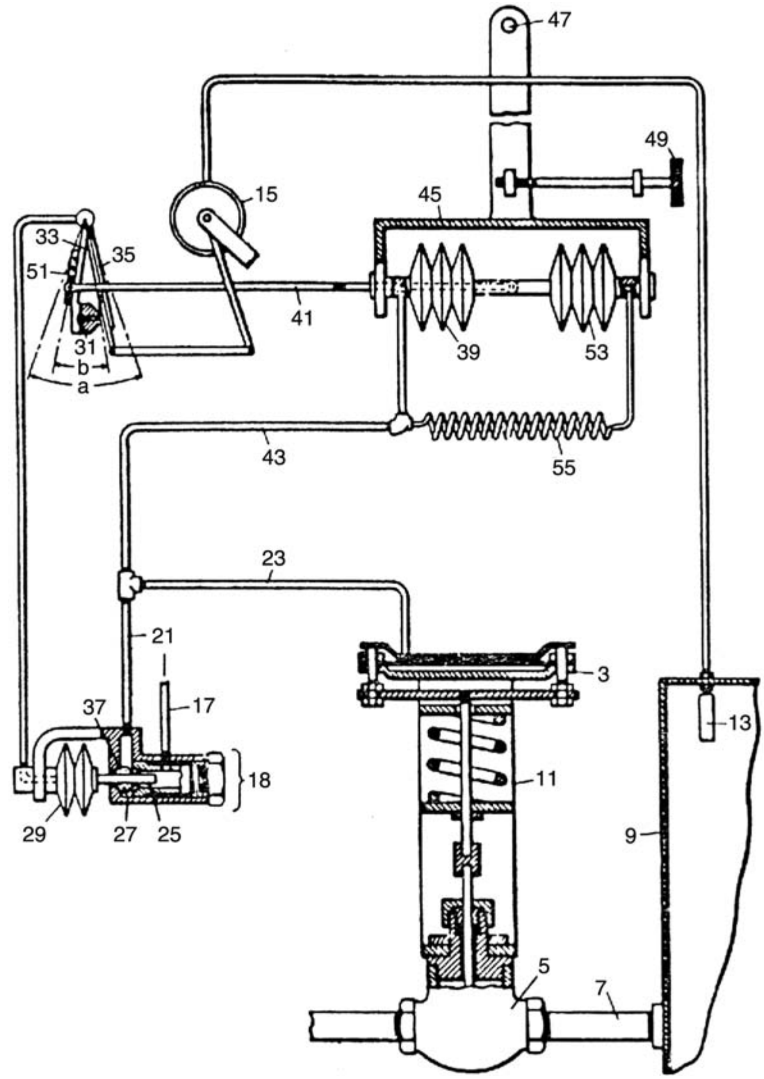


Figure 24. Basic mechanism of the Stabilog. [*Source: Bennett, S. A History of Control Engineering: 1930–1955. Institution of Electrical Engineers, London, 1993, p. 42. Reproduced with the permission of the Institution of Electrical Engineers.*]

drove the valve in a continuous fashion: the valve motion followed the time integral of the control signal. Both methods had shortcomings. The two-stage controller introduced harmful surges at the on–off transitions. The integral controller's controlled variable could not follow promptly the variation of its controlling variable—it had a time lag. In 1931, the Foxboro Company introduced a combined proportional–integral "Stabilog" controller, and the time lag was significantly reduced with the incorporation of this proportional control. In the early 1930s, Ralph Clarridge at the Taylor Instruments Company discovered that when using the time derivative of the measured variable as the control signal, the speed of reaction could further increase. In 1935, Taylor Instruments built a small experimental factory using a controller combing the proportional, integral, and derivative (PID) actions. The PID controller retained the advantages of the three individual controllers but compensated for their respective shortcomings. In fact, the PID control discovered and promoted by instrumental engineers in the 1930s was anticipated in the early 1920s by the engineer Nicholas Minorsky when he worked on theoretical problems in automatic steering. The PID scheme dominated industrial control until the 1950s.

Theoretical understanding grew with electrical control technologies. The study of telephonic repeaters produced a theory of feedback amplifiers. The development of process controllers confirmed and elaborated the theoretical work on the PID control. A unified control theory began to take shape in the late 1930s when Massachusetts Institute of Technology (MIT) researchers such as Harold Hazen, working on gunfire control systems, combined the work on feedback amplifiers, the analytic techniques involved in manufac-

turing control, and the stability analysis of power networks to form a theory of servomechanism.

See also **Fly-By-Wire Systems; Rectifiers; Smart and Biomimetic Materials**

CHEN-PANG YEANG

Further Reading

Bennett, S. *A History of Control Engineering: 1800–1930*. Institution of Electrical Engineers, London, 1979.

Bennett, S. *A History of Control Engineering: 1930–1955*. Institution of Electrical Engineers, London, 1993.

Fuller, A.T. The early development of control theory. *Trans. Am. Soc. Mech. Eng*., 98, 109–118, 1976.

Happell, G.E. and Hesselberth, W.M. *Engineering Electronics*. McGraw-Hill, New York, 1953.

Hughes, T. *Networks of Power: Electrification in Western Society 1880–1930*. Johns Hopkins University Press, Baltimore, 1983.

Mindell, D. Opening Black's box: rethinking feedback's myth of origin. *Technol. Cult*., 41, 405–434, 2000.

Stevens, R.A. Control engineering: historical background. *IEE Proc*., 121, 396, 1974.

Sze, S.M. Semiconductor Devices: Physics and Technology. Wiley, New York, 1985.

Cracking

Three major innovations emerged to meet the dramatically higher needs for quantity and better quality motor fuel in the twentieth century: thermal cracking, tetra ethyl lead, and catalytic cracking (including the fluid method). William M. Burton of Standard Oil, Indiana, in 1911–1913 raised gasoline fractions from petroleum distillation from 15 to 40 percent by increasing both temperatures and pressures. Benjamin Stillman Jr. of Yale University had discovered in 1855 that high temperatures could transform or "crack" heavy petroleum fractions into lighter or volatile components, but at atmospheric pressure over half of the raw material was lost to vaporization before the cracking temperature (about 360°C) was reached. English scientists James Dewar and Boverton Redwood discovered and patented in 1889 a process using higher pressure to restrain more of the heavier fractions and increase the volatile components. Burton and his team, not aware of this patent at the start of their work, similarly used higher pressure to improve his gasoline yields. Their process by 1913 employed a drum 9 by 2.5 meters diameter over a furnace and a long runback pipe of 300 millimeter diameter that carried vapors to condensing coils and simultaneously allowed heavier fractions to drip back into the drum for more cracking. A tank connected to the condensing coils separated gasoline from the uncondensed hydrocarbon vapors. They used a comparatively low (5.1 atmospheres) pressure because they relied on riveted plates. Burton's team improved the batch process in the next three years with false bottom plates in the still to improve cleaning time of coke deposits, and a bubble tower to improve fractionation of cracked vapors and increase gasoline yield. Refinery manager Edgar M. Clark took cracking to the next step by using tubes of about 100 millimeters diameter as the primary contact with the furnace, which increased cracking time and allowed pressures of about 6.8 atmospheres, and decreased maintenance time because of reduced coke deposition and fuel costs.

The dramatic increase in gasoline yield and the royalties that Indiana charged its licensees encouraged many to experiment with cracking. The U.S. Patent Office granted several patents for general principles or for overlapping techniques. Universal Oil Products Company (UOP) received two such patents in 1915 and 1921 incorporating the principles of continuous processing and clean circulation. Their Dubbs process (after Jesse Dubbs and his son, Carbon Petroleum Dubbs) allowed uncracked oils or reflux to be returned continuously and while still hot to the cracking zone to increase production of gasoline, and to reduce coke formation and cleaning time. The process could also charge lower value and heavier oils (such as fuel oil).

Although UOP had substantially revised the 1915 Dubbs patent from its original filing date in 1909 primarily to pursue infringement claims against the Burton process, the 1921 patent and considerable inventive work by talented personnel helped create a successful process development firm. Smaller as well as larger firms (e.g., Shell) could license the process because of its uniform royalty rate, UOP's continued technical improvements, and a solid defense against litigation from competing processes. UOP licensees and scientists further refined the Dubbs process with a hot oil pump to increase the capacity from 500 to 700 barrels per day, and contributed to the many advances in structural development, such as seamless steel tubing, electric welding, alloy steel tubing and parts, and high-temperature fans and insulation. Finally, UOP scientists like Gustav Egloff successfully urged UOP's leadership to focus on the next step of cracking—catalysis. Egloff recruited the Russian expert in catalytic reactions, Vladimir Ipatieff, who in turn brought German scientists such as Hans Tropsch to the firm.

Other cracking processes were based on broad patents having much in common with each other,

which encouraged a patent war, and then a patent pool in 1923 between four major processes. Among the major innovations were the Cross method's pressures of 600 psi (40.8 atmospheres) in 1921, achieved with a forged reaction chamber having walls 3 inches (75 millimeters) thick. This and a new centrifugal oil pump from the design of a German concern, Weiss, allowed the Cross process and Standard Oil of New Jersey to increase the cracking temperature, the rate of cracking, the throughput, and the quality of the gasoline.

Charles F. Kettering, head of DELCO and then General Motors' research division, also focused on the quality of gasoline in addressing engine knock. He and a team under his assistant, Thomas Midgley Jr., tested chemical compounds more or less systematically, and later worked with Robert Wilson of Massachusetts Institute of Technology (MIT) to compose a periodic table of more promising antiknock elements based on the spaces in their outer electron shells. In December 1921 after five months of research on organometallic compounds of the most promising elements, the team tested an obscure compound—tetraethyl lead, or TEL. Its antiknock abilities were obvious, and General Motors soon allied with its major stockholder, DuPont, to develop and manufacture the fuel additive. Standard Oil of New Jersey quickly developed a superior method to manufacture the additive, and bargained with General Motors to form Ethyl Gasoline Corporation to make and sell TEL. Poor Jersey Standard quality control led to a U.S. government investigation in 1925 of tens of lead poisoning cases in its factory, but scientific methods and equipment were not sophisticated enough to yield sufficient data to ban the material from gasoline. In attempts to sell their product to the public, Ethyl Gasoline led a drive to construct an octane rating for antiknock characteristics.

A major roadblock to increasing TEL sales was the Sun Oil Company, which already marketed a high-octane gasoline in 1927, and invested heavily in the early 1930s in developing the first catalytic cracking (Houdry) process. French industrialist and inventor Eugene Houdry had tried to interest major European oil firms in his process in the mid-1920s, but his design was not advanced enough and IG Farben's hydrogenation process appeared to be the petroleum process of the future. Houdry first worked with the Vacuum Oil Company (later Socony-Vacuum), and then interested Sun in the high octane rating (91) of his process. Sun tenaciously avoided buying the TEL additive, and plunged into developing the fixed-bed Houdry process, making significant improvements, from the design of the apparatus to the development of a synthetic catalyst. The process was semicontinuous, since oil feed would be switched from one reaction or regeneration chamber to another while the first's catalyst pellets would be cleaned of carbon. Sun's improvements increased the percentage of catalytic gasoline per charging stock from 23 to 43 percent, and by 1939 Sun and Socony had constructed or were constructing plants with a capacity of 221,675 barrels a day of charge.

The Houdry fixed-bed plants produced 90 percent of all catalytically cracked aviation gasoline in the U.S.'s first two years of World War II. Socony and the Houdry Process Corporation created a moving bed catalytic cracking (MBCC) process that continuously circulated pellets from cracking to regeneration chambers, decreased the use of steel and turbines, which were needed elsewhere in the war, reduced losses in costs of heating catalysts, and enabled the use of heavier feedstocks. Later versions dramatically increased the ratio of catalysts to charging stock, and used airlift instead of mechanical means to move the catalyst, which simplified the equipment and reduced investment costs.

Meanwhile other major firms had set about to circumvent the Houdry process and develop one of their own, to be named the fluid catalytic cracking (FCC), or fluid bed process. Jersey Standard led this endeavor with its early work and ventures with IG Farben, which had the best experience in catalytic hydrogenation for coal and petroleum by the late 1920s. Building up its catalytic research and development capabilities in the 1930s, Jersey refused to pay the steep royalty payments demanded by Eugene Houdry in 1938, and soon allied with Indiana Standard, M.W. Kellogg Company, IG Farben, British Petroleum, Shell, Texaco, and UOP in the Catalytic Research Associates to construct their own catalytic cracking process. Jersey's huge monetary and research and development resources, and its ties to top academic scientists, helped it to dominate this tremendous project.

As early as the mid-1930s Jersey had researched a process with a powdered catalyst combined with oil vapors at 400 to 540°C. Varying vapor mixes and pressures moved the mixture through a reaction chamber much like a fluid. The catalysts were separated from the vapor before it condensed and regenerated in a separate chamber. Advantages over a fixed-bed process included reduced use of steel with the separate reaction and regeneration units, a simplified continuous

process with a uniform flow and constant physical conditions, more contact between the catalyst and the oil, and the possibility of increased size to provide more economies of scale.

Later innovations, mostly by major petroleum or process design firms, include microspherical catalyst particles in 1948 to improve flow and reduce catalyst splitting and loss, high percentage alumina catalysts in 1955 to boost yield, and zeolitic catalysts in 1964 to increase yields, decrease deactivation by coke deposition, and enable shorter reaction time with lower catalyst-to-oil ratios. Zeolitic catalysts encouraged much more cracking in the transfer line (riser) before the reactor bed, while residual oil cracking became popular in the mid-1970s with increased crude oil prices and decreased demand for heavy fuel oil.

See also **Cryogenics, Applications; Feedstocks; Oil from Coal Process**

KENNETH S. MERNITZ

Further Reading

Enos, J.L. *Petroleum Progress and Profits: A History of Process Innovation*. MIT Press, Cambridge, MA, 1962.

Gibb, G.S. and Knowlton, E.H. *History of the Standard Oil Company (New Jersey): 1912–1927, The Resurgent Years*. Harper & Brothers, New York, 1956.

Giebelhaus, A.W. *Business and Government in the Oil Industry: A Case Study of Sun Oil, 1876–1945*. JAI Press, Greenwich, CT, 1980.

Larson, H., Knowlton, E.H. and Popple, C.S. *History of the Standard Oil Company (New Jersey): 1927–1950, New Horizons*. Harper & Row, New York, 1971.

Leslie, S.W. *Boss Kettering*. Columbia University Press, New York, 1983.

Mernitz, K.S. Governmental research and the corporate state: the Rittman refining process. *Technol. Cult*., 31, 83–113, 1990.

Robert, J.C. *Ethyl: A History of the Corporation and the People Who Made It*. University Press of Virginia, Charlottesville, VA, 1983.

Rosner, D. and Markowitz, G. Deadly fuel: autos, leaded gas, and the politics of science, in *True Stories from the American Past*, Graebner, W., Ed. McGraw Hill, New York, 1993, pp. 126–141.

Williamson, H.F. *et al. The American Petroleum Industry: The Age of Energy, 1899–1959*. Northwestern University Press, Evanston, IL, 1963.

Wilson, J.W. *Fluid Catalytic Cracking Technology and Operations*. Pennwell, Tulsa, OK, 1997.

Crop Protection, Spraying

Humans have controlled agricultural pests, both plants and insects, that infest crops with a variety of biological and technological methods. Modern humans developed spraying pest management techniques that were based on practical solutions to combat fungi, weeds, and insects. Ancient peoples introduced ants to orchards and fields so they could consume caterpillars preying on plants. Chinese, Sumerian, and other early farmers used chemicals such as sulfur, arsenic, and mercury as rudimentary herbicides and insecticides. These chemicals were usually applied to or dusted over roots, stems, or leaves. Seeds were often treated before being sowed.

As early as 200 BC, Cato the Censor promoted application of antipest oil sprays to protect plants in the Roman Empire. The nineteenth century potato famine and other catastrophic destruction of economically significant crops including vineyard grapes emphasized the need to improve crop protection measures. People gradually combined technological advances with biological control methods to initiate modern agricultural spraying in the late nineteenth century. Such crop protection technology was crucial in the twentieth century when large-scale commercial agriculture dominated farming to meet global demands for food. Individual farms consisted of hundreds to thousands of acres cultivated in only one or two crop types. As a result, spraying was considered essential to prevent devastating economic losses from pest damage associated with specific crops or locales.

People recognized the benefits of spraying to cover agricultural plots efficiently and evenly with protective substances and fertilizers. In the late nineteenth century, researchers developed effective inorganic chemical pesticides. Farmers initially used hand sprayers while walking through fields. By 1880, a commercial spraying machine was introduced. Four years later, the state of Utah passed a law requiring fruit trees to be sprayed.

Agricultural mechanization enabled spraying technology to advance. Machinery was used to apply spray uniformly. Automated timers regulated spraying at specified intervals. The type of machinery used depended on the type of chemical being applied. Sprays such as Monsanto's Roundup were made to be compatible with varying spraying equipment and situations. Technological advances included the design of tractors and the spray tanks that they carried or towed.

Typical sprayers distribute liquid drops forced through nozzles by pressure. The diameter of drops determines how many can be sprayed and how far they travel. A greater quantity of smaller drops can be sprayed at a time than larger drops. The smaller drops do not travel as far as the large drops because they are slowed more quickly by air resistance, but smaller drops more effectively

cover plants. Sprayers are adaptable for a variety of tasks ranging from protecting fields to individual trees.

Specialized sprayers discharge liquids with technological processes associated with atomizers, centrifuges, and nebulizers. Compared to atomizers, sprayers require less energy and are inexpensive. Sprayers' drop size, however, necessitates more liquid to spray than atomizers and are not as comprehensive in their coverage. Fans in atomizers create an air jet, which shatters liquid drops into smaller droplets that spread over a larger surface than drops from sprayers. Atomizers require less time than sprayers to treat an area and their increased plant coverage improves chemical performance. The lightweight droplets are sometimes diverted by wind but are more likely to be absorbed by soil than to runoff, as is the case with spray droplets. They are, however, prone to rapid evaporation. Atomizers cost more and need more energy than sprayers. They are often worn like backpacks and equipped with wands to spray.

In contrast, nebulizers deliver drops with a fog of steam or gas that envelops plants. Nebulizer fogs are described as coarse or wet, or fine or dry depending on drop size. These devices are usually found in greenhouses because the tiny particles are vulnerable to weather conditions. Nebulizers are more effective than other forms of spraying but are expensive to buy and maintain. Most centrifugal sprayers are run manually and have a small electric motor powered by batteries. Gravity pulls liquids in centrifugal sprayers through rotary disks to create small particles.

Small airplanes were first utilized for crop spraying in the 1920s. Many pioneering crop dusters were World War I veteran pilots. Aviation enables large areas to be treated in a brief time compared to hand or machinery-pulled spraying tools, and isolated places could be easily reached by air. In the southern U.S., airplanes were used to spray calcium arsenate over cotton fields in an attempt to destroy boll weevils which threatened to ruin the important cash crop. Although more expensive, helicopters can reach places that planes cannot go and their rotating blades push chemical sprays toward their targets. Wind limits all agricultural aircraft spraying applications.

DDT (dichloro-diphenyl-trichloroethane) was identified as an effective insecticide in 1939. During the early 1940s, DDT transformed crop-spraying protection. Researchers addressed the increase of pest resistance to chemicals, especially DDT, in the 1950s and 1960s. Genetic engineering efforts to produce transgenic plants resistant to pests and diseases were expanded as a supplement or alternative to chemical sprays.

Some people recognized risks associated with spraying. In 1892, a Canadian law prohibited the chemical spraying of blooming trees in order to protect bees collecting pollen. Twentieth century critics of spraying stressed the expensive costs of sprays, possible health risks to agriculturists and consumers, and the threat of environmental damage and pollution. They warned of agriculturists becoming too dependent on sprays. When spray schedules were disrupted or cancelled, pests became more problematic. Scientists sought alternative and less hazardous methods. The British Royal Commission on Arsenical Poisoning established residue limits in 1903. Twenty years later, the British government threatened to refuse importation of U.S. apples that had been exposed to chemical sprays. The U.S. Food and Drug Administration devised spraying standards for exported fruit.

Rachel Carson outlined her concerns regarding pesticides applied by spraying in *Silent Spring* (1962). Some scientists also voiced ecological concerns related to sprayed chemicals that protected crops. R.F. Smith and Robert van den Bosch devised the concept of the more environmentally sound integrated pest management (IPM) in 1967. As a result of the expanding environmental movement, DDT was banned in many places worldwide in the 1970s. Several laws regulating spraying, including licensing users, were enacted. Such official recognition continued into the 1990s when more pathogenic species of insects, weeds, and animals became resistant to chemical sprays and new spraying formulas and techniques were tested.

See also **Farming, Agricultural Methods; Genetic Engineering, Applications; Pesticides**

ELIZABETH D. SCHAFER

Further Reading

Anderson, M.I. *Low & Slow: An Insider's History of Agricultural Aviation*. California Farmer Publishing, San Francisco, 1986.

Cate, J.R. and Hinkle, M.K. *Integrated Pest Management: The Path of a Paradigm*. National Audubon Society, Washington D.C., 1994.

McWhorter, C.G. and Gebhardt, M.R., Eds. *Methods of Applying Herbicides*. Weed Science Society of America, Champaign, Illinois, 1987.

Rasmussen, H. *Crop Dusters: "Props in the Crops."* Osceola, WI, Motorbooks International, 1986.

Van den Bosch, R. *The Pesticide Conspiracy*. University of California Press, Berkeley, 1978.

Wilkins, R.M., Ed. *Controlled Delivery of Crop-Protection Agents*. Taylor & Francis, London, 1990.
Yuste, M.-P. and Gostincar, J. *Handbook of Agriculture*. Marcel Dekker, New York, 1999.

Cryogenics, Applications

The field of cryogenics, and its applications, has evolved from intertwined interests of the nineteenth century—the scientific interest to liquefy the remaining ''permanent'' gases, the residential interest for indoor environmental control, and the commercial interest in food preservation. A rich diversity of applications has resulted, ranging in daily impact and technical sophistication from frozen foods to rocket science. A survey of this variety is summarized in the following paragraphs.

Biomedical Applications

Although the medical field has benefited indirectly from cryogenics through the superconducting imaging technologies of magnetic resonance imaging (MRI) and SQUID detectors, direct benefits arise through various forms of cryosurgery and transport of biologicals. Significant use of a liquid nitrogen (LN_2)-based cryosurgical probe was initiated in the mid-1960s for treating brain tissue to alleviate the effects of Parkinson's disease. Throughout the century dermatologists have used a spray or swab application of LN_2 to destroy undesirable tissue. In the 1980s physicians were able to precisely treat tumors in the liver and prostate with cryosurgical devices and the the emerging imaging technologies of ultrasound, computer tomography, and MRI. In the 1990s cooling for the cryoprobes became increasingly convenient as the integration of miniature and micro cryocooler technologies replaced the use of LN_2. A significant benefit of the cryosurgical devices is the ability through partial, reversible cooling to investigate the efficacy of freezing tissue areas before permanent treatment. Additional treatments are emerging for breast cancer, benign tissue in gynecology, heart arrhythmia, and others.

Improvements to cryogenic storage vessels through the twentieth century have enabled a wide variety of biological samples to be preserved. Beginning in 1956, cryopreservation of bull semen spread the practice of scientific breeding to developing nations, resulting in a significant addition to world supplies of milk and meat. Transport dewars have also been used for storing blood, corneas, virus samples, biopsy specimens, and heart valves.

Cryocoolers

In 1956 J.W.L. Kölher at the Philips Company in Holland developed the first commercial Stirling refrigerator, by reversing the Stirling cycle previously used for power generation. Beginning in the 1960s this compact, closed-cycle cryogenic refrigerator, or cryocooler, has been utilized extensively for infrared detection in security systems, astronomy, and in military applications for night vision goggles and heat-seeking devices. A family of commercial cryocoolers, differing in their thermodynamic cycles, have been developed through the latter half of the twentieth century, including the Gifford McMahon, Vuilleumier, and miniature versions of Joule–Thomson and Reverse-Brayton cycles in the 1960s and 1970s, and the orifice pulse tube in the 1980s and 1990s. The advantages of compactness and ability to create cryogenic temperatures without liquid cryogens have resulted in continually expanding applications of cryocoolers. Examples include the over $200 million per year cryopump industry, cold electronics, fetal heart monitors, and remote-site liquefaction of air, nitrogen and oxygen.

Food Processing

Beginning in 1923 with Birdseye's processing plant for frozen fish, the food industry has capitalized on the benefits of rapid freezing afforded through the use of cryogenics. Additional benefits such as retained texture, color, and flavor have maintained the growth of this industry throughout the twentieth century.

Materials

During the latter half of the twentieth century a cryogenic process was developed to improve the strength and wear-resistance of certain metals. Tests in 1944 in which alloy steel was cold-cycled in LN_2 resulted in minor improvements to wear resistance, but also increased brittleness. In the early 1970s a different approach tested by the Advanced Engineering Metal Parts Division of the Sperry Rand Corporation revealed 150 to 300 percent increased life for machining tools. R.F. Barron at Louisiana Tech University extended the investigation, determining an optimum process for high-speed lathe tools and demonstrating improvements in wear resistance by factors of 2 to 5 with five different tool steels. The optimized process involves very slow cooling (2°C per hour) to 80°Kelvin, followed by a 1- to 2-hour soak in LN_2. Although the warm-up rate to ambient conditions is not crucial, a mild temperature at

150°C for 1 hour is important for eliminating brittleness. The process aids the conversion in steels from a large-grain austenitic microstructure to the denser, finer grain-sized martensitic microstructure with increased hardness and wear resistance. Other applications of this technology emerged in the 1990s with the establishment of more than 50 small-scale companies treating gun barrels, musical instruments, guitar strings, and engine parts for motorcycles and racecar engines.

Space

Space exploration has been distinctively enabled by the cryogenics industry. The availability of large quantities of liquid hydrogen and liquid oxygen facilitated the development, from the 1950s onward, of rockets and other space vehicles. Space technology now utilizes most of the cryogens for propulsion, sensor cooling, and life support. The space applications of helium, beyond pressurizing other cryogen tanks, are especially unique. The unusual properties of superfluid helium (liquid helium below 2.17°Kelvin) such as zero viscosity, unusually high heat transport, and a thermomechanical pressure, first explored during the 1930s, have been ingeniously utilized where zero gravity necessitates special requirements for coolant transfer and containment. Helium refrigeration technologies have enabled space-based infrared detectors to operate at 60 millikelvin.

Sub-Kelvin Technology

Liquid helium has enabled research at very low temperatures primarily through the use of two different refrigeration techniques. The adiabatic demagnetization technique, developed by Giauque and MacDougall in 1933, utilizes cooling provided by the magnetocaloric effect in certain paramagnetic salts to produce temperatures down to 1 millikelvin. The dilution refrigerator, proposed by H. London in 1951, and first developed at the University of Leiden by Das, Ouboter, and Taconis in 1965, evaporates pure liquid He^3 into a dilute mixture with superfluid He^4 to produce temperatures down to 2 millikelvin. He^3 is the lighter isotope (2 protons, 1 neutron) of helium. From the mid-1960s to the mid-1990s dilution refrigerators became the preferred method of reaching sub-Kelvin temperatures. However, with the growing convenience of superconducting magnets, compact adiabatic demagnetization devices are again becoming popular. The introduction and growth of superconducting magnets, enabled by ready supply of liquid helium, is described elsewhere.

See also **Cryogenics, Liquifaction of Gases; Food Preservation: Cooling and Freezing; Superconductivity, Applications; Tomography in Medicine**

JOHN M. PFOTENHAUER

Further Reading

Barron, R.F. *Cryogenic Systems*. Oxford University Press, New York, 1985.

Baust, J.G., Gage, A.A., Ma, H. and Zhang, C.M. Minimally invasive cryosurgery-technological advances. *Cryobiology*, 34, 373–84, 1997.

DeGaspari, J. The big chill. *Mech. Eng*., 122, 11, 94–97, 2000.

McGrath *et al*. Cryobiology update. *Cold Facts*, 15, 3, 1–26, 1999.

Miller, J.P. The use of liquid nitrogen in food freezing, in *Food Freezing: Today and Tomorrow*. Springer Verlag, Berlin, 1991.

Mendelssohn, K. *The Quest for Absolute Zero: the Meaning of Low Temperature Physics*. Taylor & Francis, London, 1977.

Shachtman, T. *Absolute Zero and the Conquest of Cold*. Houghton Mifflin, Boston, 1999.

Timmerhaus, K.D., Fast, R.W. and Kittel, P., Eds. *Advances in Cryogenic Engineering*. Plenum Press, New York, 1960.

Walker, G. *Cryocoolers: Part 1 Fundamentals; Part 2 Applications*. Plenum Press, New York, 1983.

Walker, G. *Miniature Refrigerators for Cryogenic Sensors and Cold Electronics*. Oxford University Press, New York, 1989.

Cryogenics, Liquefaction of Gases

The modern air liquefaction industry generates annual sales in the tens of billions of dollars through production of liquefied oxygen, nitrogen, argon, and specialty gases. Following liquefaction, air is fractionally distilled to separate the various components. The cryogenic liquids are then stored and transported in specially insulated vessels and shipped to thousands of customers for a variety of applications.

Oxygen is utilized in the manufacture of steel, other metals, and glass, in the pulp and paper industry, in chemical processes, and for life support in medical applications. Nitrogen finds application in chemical processes such as the production of ammonia, in rapid-freeze food processing, in biomedical systems that preserve blood, tissue or semen, and in ecological recycling of car tires. Argon provides an inert atmosphere for the manufacturing of steel and aluminum, for welding, and for fabricating electronics components. It is

also used for lighting, as are neon, krypton, and xenon. The two coldest cryogenic liquids, hydrogen and helium, are utilized in various industries, hydrogen finding uses in chemical and food processing, pharmaceuticals, fuel propellant for spacecraft, and in emerging fuel cell technologies. Helium is used extensively in the aerospace industry, as an inert atmosphere in metal processing, for leak detection in HVAC and vacuum systems, and as a coolant for superconducting magnets and low temperature research.

Producing and using the liquid cryogens has been a leapfrog process in which the possibility of obtaining the pure gases has spawned new applications, with the resulting applications motivating new possibilities in the cryogenics industry. With the exception of utilizing oxygen for making steel, and nitrogen for agricultural fertilizer, none of the applications cited above existed in 1900. It was only in 1895 that the patents enabling commercial production of liquid air, oxygen, and nitrogen, were issued to Carl von Linde in Germany and William Hampson in Great Britain. Figure 25 schematically diagrams the mechanical components and flow of gas through the Linde–Hampson air liquefier. As the air is compressed it rejects heat to its surroundings. Additional cooling of the pressurized air occurs in the counter-flow heat exchanger from the colder return-gas. Exiting the heat exchanger, the pressurized air is expanded through a valve resulting in a drop in pressure and temperature, to the extent that some of the air exits the valve in liquid form. The remaining cold gas returns up through the counterflow heat exchanger. Being warmed by the incoming pressurized air, it returns to the suction side of the compressor near room temperature. Adding the same mass in gaseous form at the inlet of the compressor then compensates for the amount of liquid that is extracted from the machine.

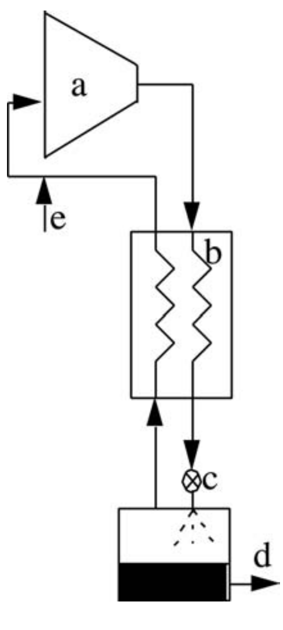


Figure 25. Linde–Hampson air liquefier: (a) compressor; (b) counterflow heat exchanger; (c) expansion valve; (d) liquid air outlet; and (e) make-up air inlet.

In 1902 Georges Claude improved the design of air liquefiers by replacing the expansion valve with an expansion piston, producing liquid more efficiently because the gas is cooled by "doing work" as it pushes against the piston during the pressure drop. In the same year, and because of the increased availability of oxygen, oxyacetylene torches replaced less effective air-acetylene torches used in welding. Although this advance was enjoyed only in Europe up to 1906, it spread along with the air liquefaction industry to other continents and is still in widespread use today.

The British Oxygen Company introduced the air liquefaction process to the U.S. at the world's fair in St. Louis in 1904. By January of 1907, the Linde Air Products Company was established in Buffalo, New York. Within 14 years the worldwide production of liquid oxygen doubled. To reduce distribution costs, cryogenic vessels holding liquefied gases increasingly replaced gas cylinders. However, inadequate insulation in the cryogenic vessels prohibited distribution over the large distances encountered in the U.S. Here the applications necessitated improvements to the liquid cryogen industry. In 1900, the best thermally insulating containers were those invented in 1892 by Sir James Dewar. These laboratory-based double-walled glass vessels used a high degree of vacuum in the space between the walls to achieve their insulating properties, but were impractical for industrial purposes. The "powder in vacuum" design, developed and commercialized by Leo Dana at the Linde Division of Union Carbide in the 1930s and 1940s, provided adequate insulation while reducing vacuum requirements and improving constructability. In this approach, the double-wall space of metal containers is filled with carefully selected powders. Containers utilizing this design today transport liquid oxygen, nitrogen, and argon worldwide for periods of several weeks without significant evaporation.

The 1940s brought the first commercial liquefaction of natural gas to improve the distribution of the popular fuel to homes and industries. Natural gas, used from the mid-1800s, is obtained

from underground cavities and is primarily composed of methane. The 600-fold increase in the density of liquid natural gas (LNG) as compared to the gas at ambient conditions, motivated the development of the LNG production facilities. The first commercial LNG plant was installed in Cleveland, Ohio in 1941 and liquefied 200 cubic meters per day. A storage tank failure in 1944 and subsequent fire caused serious damage and loss of lives, delaying growth of the industry for over ten years. Subsequent installations, with improved safety, were established first as a floating barge in 1956, and later as permanent land-sites worldwide. Today, the multibillion dollar LNG industry transports LNG via 41,000-liter-tank trucks and 125 000-cubic meter-capacity ships.

While the liquid air industry was beginning in the U.S., the scientific race to condense the remaining unliquefied gas approached its finish line. On 10 July 1908. Kamerlingh Onnes at the University of Leiden succeeded in liquefying helium at 4.2°Kelvin. Liquid helium remained a laboratory curiosity until the invention of a convenient helium liquefier by Sam Collins of the Massachusetts Institute of Technology (MIT) in 1947. In 1952 the National Bureau of Standards in Boulder, Colorado was established to develop a large-scale hydrogen liquefier. By the 1960s liquid helium and hydrogen were being used extensively for emerging programs in space exploration and superconductivity. Union Carbide's development of multilayer insulation (MLI) or superinsulation enabled the 100-fold improvement to the cryogenic storage vessels that allowed large-scale transport of liquid helium and liquid hydrogen.

See also **Cryogenics, Applications**

JOHN M. PFOTENHAUER

Further Reading

Dana, L.I. with Donnelly, R.J. and Francis, A.W., Eds. *Cryogenic Science and Technology: Contributions by Leo I. Dana*. Union Carbide, 1985.

Isalski, W.H. *Separation of Gases*. Oxford Science, Oxford, 1989.

Mendelssohn, K. *The Quest for Absolute Zero: the Meaning of Low Temperature Physics*. Taylor & Francis, London, 1977.

Shachtman, T. *Absolute Zero and the Conquest of Cold*. Houghton Mifflin, Boston, 1999.

Scott, R.B. *Cryogenic Engineering*. Met-Chem Research, Boulder, CO, 1988.

Scurlock, R.G. A brief history of cryogenics, in *Advances in Cryogenic Engineering*, 37A, Fast, R.W., Ed. Plenum Press, New York, 1991.

Timmerhaus, K.D. and Flynn, T.M., *Cryogenic Process Engineering*. Plenum Press, New York, 1989.

Williamson Jr., K.D. and Edeskuty, F.J. *Liquid Cryogens: Volume I Theory and Equipment, Volume II Properties and Applications*. CRC Press, Boca Raton, FL, 1983.

Crystal Detectors, *see* **Radio Receivers, Crystal Detectors and Receivers**

Crystals, Synthetic

A crystal is an immobilized arrangement of atoms, ions or molecules packed in a regular array. Although naturally occurring crystals can be things of beauty, it is rare to find a fault-free specimen of any considerable size. Crystals can now be synthesized as gems, but more often they are manufactured for utility. This discussion on synthetic crystals will outline some of the techniques that are used in the manufacture of technologically important crystalline materials other than semiconductors, which are treated elsewhere in this encyclopedia.

One of the first applications of crystalline materials was in specialist optics. In 1828 William Nicol of Edinburgh developed a prism based on calcite ($CaCO_3$) crystals which could polarize light. Two such prisms were used in a polaroscope (an instrument for measuring the concentration of optically active chemicals) and the design remained unchanged until the 1980s. At first, piezoelectricity (a voltage causing mechanical deformation and vice versa in a crystal) was a curiosity but it assumed critical importance in 1917 when Paul Langevin assembled a mosaic of thin quartz crystals as a passive sonar array for submarine detection. Quartz crystals became very important as radio engineers struggled to "lock" the frequencies of transmitters so that more users could be accommodated in the limited radio-frequency spectrum.

We can assume that by the end of World War I, the demand for certain crystals outstripped the supply of suitable naturally occurring materials. Fortunately, techniques for manufacture were already to hand and others have been developed throughout the twentieth century. These have been driven by technological advances, such as the requirements of the semiconductor industry. The demand for synthetic ruby grew enormously after it was used in the first lasers in 1960.

For the purposes of classification the plethora of techniques can be listed as: growth from a melt, growth from solution, growth involving high temperature, growth involving high pressure, and growth involving both high temperature and pressure. However, before discussing these we should

remember that nature does not like a state of order and that is exactly what a crystal is. Therefore, crystals will only grow if there is a net benefit, such as a significant decrease in energy. The physical chemistry of the crystal formation is concerned with phase diagrams and with equilibria. The latter dictates that growth must normally be a slow process if defect-free crystals are to be obtained.

Although the French chemist, Edmond Fremy, had been producing commercial quality gemstones in 1877, the first major advance was the production of ruby using the Verneuil process (developed by French chemist Auguste Victor Louis Verneuil) in 1902. In this technique a seed crystal is held under an oxy-hydrogen flame and the base material (alumina, Al_2O_3 powder with an addition of chromia, Cr_2O_3 as a contaminant to give color) is fed through the flame. Other materials such as sapphire, rutile (titanium dioxide, a gem that is also used as a white pigment in paints, paper, plastics, sunscreens, and cosmetics), spinel (a mineral that can be used to imitate blue sapphire or aquamarine; also has good optical qualities for laser applications), and strontium titanate (has no natural counterpart, but can imitate diamond; and as a ceramic has been widely used for various electronic applications) were later made using this method. Synthetic gems have the same chemical composition and crystal structure as natural gems, but lack irregularities or tiny inclusion imperfections that give real gems flaws that produce their unique appeal.

The Czochralski method published by Jan Czochralski in 1918 involves dipping a seed crystal into a container of molten base material. The seed is rotated as it is slowly withdrawn and the material adhering to it assumes the same crystal orientation. This process has been used to make ruby, sapphire, spinel, and yttrium-aluminum-garnet (YAG), an important material for lasers. The Czochralski technique was later to become the cornerstone of the silicon industry and bars of single-crystal silicon up to 250 millimeters in diameter are now in regular production. Czochralski growth of III–V semiconductors such as gallium phosphide presents considerable problems because of the volatility of phosphorous. This has been overcome by adding boron oxide to the melt; which is immiscible and floats on top, rather like oil on water, inhibiting evaporation.

Growth from aqueous solution is historically much older than any of the other techniques, but it remains very important in the manufacture of certain crystals; for example, Rochelle Salt (potassium-sodium tartrate) is a much more effective piezoelectric material than quartz and in 1917 Alexander Nicolson at Bell Laboratories demonstrated that it could be used for sonar applications. In single-crystal form, potassium dihydrogen phosphate (KDP) is a nonlinear optical material used for doubling, tripling and quadrupling the frequency of the output from high-power lasers. In 1998 the world's largest KDP crystal (250 kilograms) was grown at the Lawrence Livermore National Laboratory for use in the Megajoule Laser Fusion Experiment.

The same growth techniques can be applied in nonaqueous systems. In the particular case of growth of crystals of materials that have an extremely high melting point the solvent is chosen for its flux properties; that is, its ability to induce growth at significantly lower temperatures. Thus aluminum oxide crystals have been grown in a flux of lead fluoride at 840°C. Once the system has cooled down the flux is dissolved in nitric acid to reveal alumina crystals.

Aqueous growth under high pressure (hydrothermal growth) is a useful means of growing otherwise difficult crystals. It is currently the preferred technique for the production of synthetic quartz.

There have been many attempts to produce synthetic diamond but the crystallization of carbon requires both high temperature and high pressure. It was not until materials such as tungsten carbide became available in 1930 that work could commence on high-pressure containment systems. The use of self-sealing techniques, such as that developed by Percy W. Bridgman, were also essential. Diamonds were produced by Baltzar von Platen and colleagues at the ASEA Laboratory in Stockholm in 1953, but the work was kept secret. Thus in 1955 Francis Bundy, Tracy Hall, Herbert Strong and Robert Wentorf at General Electric were able to claim the first commercial production of synthetic diamond. Today, films of diamond can be produced using less severe conditions. This has developed as an offshoot of a technique called vapor phase epitaxy, which is very important in silicon device manufacture and involves the growth of crystalline layers by chemical vapor deposition (CVD). For example gaseous silane (SiH_4) or silicon trichoride ($SiHCl_3$) is carried through a reaction chamber in a stream of hydrogen. At the center of the reactor is a graphite block held at approximately 1200°C with one or more silicon substrates resting on top. The silane breaks down on the hot surface to yield silicon and hydrogen and the deposited material has the same the crystal orientation as the substrate,

This summary lists some of the methods for growing technologically important crystals. Improvements continue, but total innovations are rare. One of the few exceptions has been molecular beam epitaxy (MBE), where crystals are grown under extreme high vacuum conditions. It is costly, but yields tailor-made materials called "superlattice crystals" which have considerable promise for the future. Perhaps the next step is routine crystal fabrication under zero gravity conditions in space.

See also **Semiconductors, Crystal Growing and Purification; Thin Film Materials and Technology**

DONARD DE COGAN

Further Reading

Gilman, J.J. *The Art and Science of Growing Crystals*. Wiley, New York, 1963.

Hazen, R.M. *The New Alchemists: Breaking Through the Barriers of High Pressure*. Times Books, Random House, New York, 1993.

Hazen, R.M. *The Diamond Makers*. Cambridge University Press, Cambridge, 1999.

Hughes, R. Synthetic corundum, in *Rubies and sapphire*. RWH Publishing, Boulder, CO, 1998.

Scheel, H. J. Historical introduction, in *Handbook of Crystal Growth*, Hurle, D.T.J., Ed., vol. 1. Elsevier, Amsterdam, 1993.

Scheel, H.J. and Fukuda, T., Eds. *Crystal Growth Technology*. Wiley, New York, 2003.

D

Dairy Farming

Throughout the world, especially in the Northern Hemisphere, milk, cheese, butter, ice cream, and other dairy products, have been central elements of food production. Over the centuries improvements in cattle breeding and nutrition, as well as new dairy techniques, led to the increased production of dairy goods. Hand-operated churns and separators were used to make butter and cream, and those close to a barnyard had access to fresh milk.

By the late nineteenth century, new science and technology had begun to transform dairy production, particularly in the U.S. and Europe. Rail transportation and iced and refrigerated boxcars made it easier to transport milk to more distant markets. Successful machinery for separating milk from cream came from the DeLaval Corporation in 1879, and the Babcock butterfat tester appeared in 1890. The first practical automated milking machines and commercial pasteurization machines were in use in the decades before 1900. Louis Pasteur's contribution to the dairy industry—discovering the sterilization process for milk—was substantial. By heating milk, pasteurization destroys bacteria that may be harmful to humans. The pasteurization process also increases the shelf life of the product by eliminating enzymes and bacteria that cause milk to spoil. Milk is pasteurized via the "batch" method, in which a jacketed vat is surrounded by heated coils. The vat is agitated while heated, which adds qualities to the product that also make it useful for making ice cream. With the "continuous" method of pasteurization, time and energy are conserved by continuously processing milk as a high temperature using a steel-plated heat exchanger, heated by steam or hot water. Ultra-high temperature pasteurization was first used in 1948.

Key to the mechanization of the dairy industry were improvements in the mechanical milking machine. The first vacuum pump machines were patented in the 1860s, and the Mehring hand- and foot-powered model was popular in the 1890s. This machine was powered by a person sitting down, pushing pedals with their feet, and milking two cows at one time. Widespread adoption of the milking machine did not come until the introduction of the DeLaval milker in 1918. Using these early machines, one person could milk 30 cows twice per day; with improvements in milking machines by the end of the twentieth century, over 200 cows could be milked by one person. What had been seen as "women's work," the tending of dairy herds, had become a business process in which farm factories produced raw milk for shipment into central dairies for processing. Elaborate dairy barns gave way to high-tech milking parlors where cows could be milked by machine three times per day rather than once or twice, and a premium was placed on sanitary conditions. By the late twentieth century, it was not uncommon in California to see dairy operations numbering up to 50,000 animals. Numerous devices were developed to improve mechanical milking, and the intermittent "pulsator"-type milker was one of the more important innovations. Finding a device that could be kept sanitary, that would maximize production per cow and that would not harm the animal's teats, were all obstacles in developing a practical milking machine. The "thistle"-type machine, so called because of the appearance of the equipment that fit on the cow's udders, was developed in the 1910s

and was the basis for the design of all subsequent milking machines. In the late twentieth century machines, the cow's teats are inserted into rubber suction cups, which are surrounded by stainless steel cups. The vacuum of the automatic machine inflates and deflates the rubber cup, with the milk flowing onto a central collection unit. Modern dairy operations are designed to be easily cleaned, and the emphasis on health and safety with the homogenization and pasteurization processes helped to increase consumer confidence in the nutritional value of dairy products.

In addition to the mass production of flavor-controlled, vitamin-fortified homogenized milk by 1919, new delivery technologies made dairy products more accessible to the public. By 1938 bulk tank trucks with coolers transported milk from dairies as dairy operations grew larger and on-farm storage demands increased. Milk bottle fillers in the 1940s could fill 4500 liters per hour, compared with earlier models that bottled 1100 to 1500 liters per hour. Plastic-coated paper containers were introduced in the 1930s, and the first all-plastic milk containers appeared in stores in 1964. By then machines could fill 23,000 liters per hour.

Increased production of high quality dairy products and better distribution of them resulted in greater consumption of dairy products in the twentieth century. The United States Department of Agriculture (USDA) created the Department of Agrostology in 1895 to study how different grasses and feedstuffs affected milk production and milk quality. The USDA's Bureau of Dairy Industry fostered better herd management and breeding techniques, more efficient and sanitary production, and new and better types of dairy products for the consumer. Agricultural researchers contributed to the dairy industry with herd improvements through artificial insemination of cattle (1938) and embryo transfer (1980). Scientists also provided a variety of antibiotics including penicillin to control mastitis and other frequent dairy diseases. Genetic modification to enhance production also came into play with the introduction of recombinant bovine somatotrophin (rBST) in 1994. That same year the U.S. Nutrition Labeling and Education Act became law, requiring producers to list rBST on the label if it is in their milk.

At the end of the twentieth century northern Europe and North America produced three-fourths of the world's raw milk, with India and South America the major secondary producers. While low-fat dairy products became increasingly popular by the 1990s, per capita ice cream consumption in the U.S. increased to 7.5 kilograms per year. In the state of California, not traditionally associated with dairy production, 20 percent of U.S. milk production came from 2,200 dairies that produced some 16 billion kilograms of raw milk.

See also **Farming, Agricultural Methods; Farming, Growth Promotion; Farming, Mechanization**

RANDAL BEEMAN

Further Reading

El-Hai, J. *Celebrating Tradition, Building the Future: Seventy-Five Years of Land O' Lakes*. Land O' Lakes, Minneapolis, 1996.

Hurt, R.D. *American Agriculture: A Brief History*. Iowa State University Press, Ames, 1994.

Moore, E.G. *The Agricultural Research Service*. Praeger, Washington D.C., 1967.

Porter, A.R. *et al. Dairy Cattle in American Agriculture*. Iowa State University Press, Ames, 1965.

Weimar, M.D. and Blayney, D.P. *Landmarks in the US Dairy Industry*. U.S. Department of Agriculture, Washington, D.C., 1984.

Dams

For thousands of years, dam and water storage technologies have allowed civilizations to flourish in parts of the world where dry climates would otherwise limit human settlement. As early as 3000 BC, civilizations along the Tigris, Euphrates, Ganges, and Nile Rivers constructed earth and stone dams across these large rivers. These structures allowed them to store water for agriculture and create complex societies on that basis.

A dam consists of a mass of earth, timber, rock, concrete, or any combination of these materials that obstructs the flow of water. A dam can either divert water or store it in a reservoir, the artificial body of water that a dam creates. Diversion dams (weirs) raise the elevation of a river and divert water into a canal for transport to a mill, power plant, or irrigated field. Storage dams impound water in a reservoir.

There are three major types of dams—gravity, arch, and buttress. Gravity dams rely for stability on their weight to resist the hydrostatic, or water, pressure exerted by the reservoir. Arch dams, built along arcs that curve upstream into reservoirs, are most commonly found in narrow canyons with hard rock foundations. The arch dam transmits the horizontal water thrust to the abutments. Multiple-arch dams consist of a number of single arches supported by buttresses. Like gravity dams, buttress dams rely on gravity for stability, but require less material than standard gravity structures. They resist hydrostatic loads by using the same engineer-

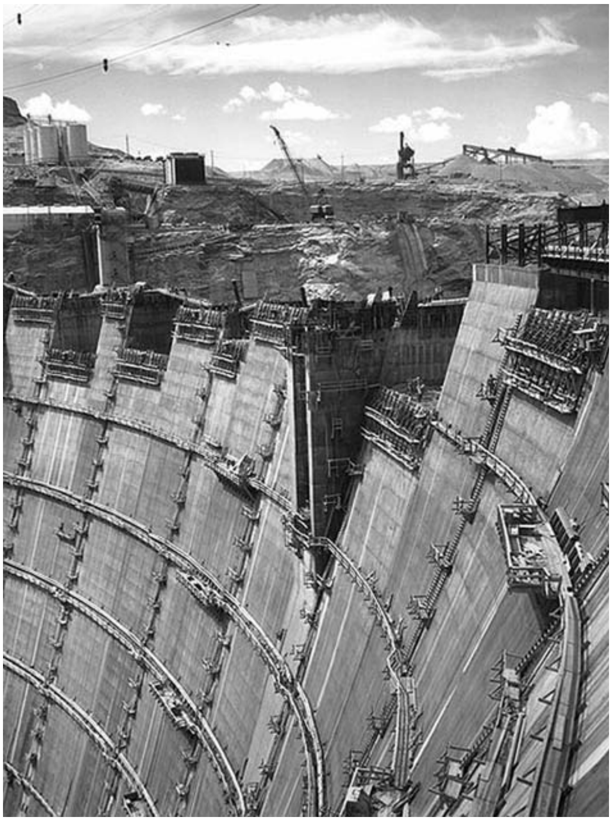


Figure 1. Construction at Glen Canyon Dam, Colorado River Basin Storage Project, 1963.
[*Photograph courtesy of the Department of the Interior, U.S. Bureau of Reclamation*].

ing principles of the flying buttresses that braced the high walls of Gothic cathedrals.

Engineers of the late Roman Empire built the first known arch dams and several buttress dams. In medieval times, Moorish engineers harnessed water from the Sierra Nevada and constructed massive irrigation systems in southern Spain. Persia's Ilkhanid Mongols built the world's tallest arch dams at that time as well. In the sixteenth century, Spanish engineers initiated a new era of dam building by applying mathematical principles to dam design. From the late eighteenth to early nineteenth centuries, French and English engineers, including Bernard Belidor, Charles Bossut, Charles Augustin Coulomb, and Henry Moseley, designed dams further based on mathematical principles. By the end of the 1800s, Europeans had initiated large dam projects around the world, particularly in areas of British control. British engineers completed Egypt's Aswan Dam in 1902, enlarged it in 1934, and rebuilt it upstream in 1960. The dam resulted in increased agricultural and industrial capacity for the region's growing population.

The twentieth century heralded the development of large, multipurpose water projects that could control floods, store water for irrigation, municipal, and industrial use, improve the navigability of waterways, and provide water for the generation of hydroelectric power. Indeed, one of the century's great advances came with the development of technologies that could generate electricity from the motion of falling water. The introduction of concrete as a construction material allowed dam designers from the 1910s to the 1930s to consider thin shells and complex curved shapes to minimize the volume of concrete required and the overall cost of arch dams. The face of a double-curvature (dome) arch dam is curved in two directions: from side to side, transferring force into the canyon walls, and from top to bottom, transferring force into the canyon floor. Posttensioned steel rods or cables commonly reinforce concrete arch dams and gravity dams, reducing the cross-section (and volume of materials) for a conventional gravity dam of the same height. Vertical steel rods, stressed by jacks and securely anchored into the rock foundation, resist the tendency of the thin shells to overturn.

The U.S. Bureau of Reclamation's Hoover Dam (originally called Boulder Dam) (a curved gravity dam, see Figure 2) served as an early prototype for the century's multipurpose water projects. Built along the Colorado River in 1935, Hoover Dam stands nearly 222 meters high—the tallest dam in the world at that time. The dam created Lake Mead, the largest man-made lake in the U.S. Hoover Dam Power Plant's 17 large turbines generate a total capacity of 2,074,000 kilowatts of electrical power for Nevada, Arizona, and California.

Hoover Dam served as a model for multipurpose water projects around the world, including government projects in Pakistan and India. Other examples of major late twentieth century dam projects include Syncrude Tailings Dam in Alberta, Canada; Rogun Dam in Tajikistan; Tehri Dam in the Indian Himalayas; and the Itaipú Dam on the Paraná River between Brazil and Paraguay. By harnessing large rivers' water and energy, these dams allow for massive agricultural and industrial development in areas with natural limitations.

Perhaps the twentieth century's most ambitious dam project was Three Gorges Dam in China, the largest construction scheme in China since the Great Wall. Started in 1992 and completed through the turn of the twentieth century, the dam was built in order to rid the fabled Yangtze River of its deadly floods, provide an important source of electrical energy for China, and open up

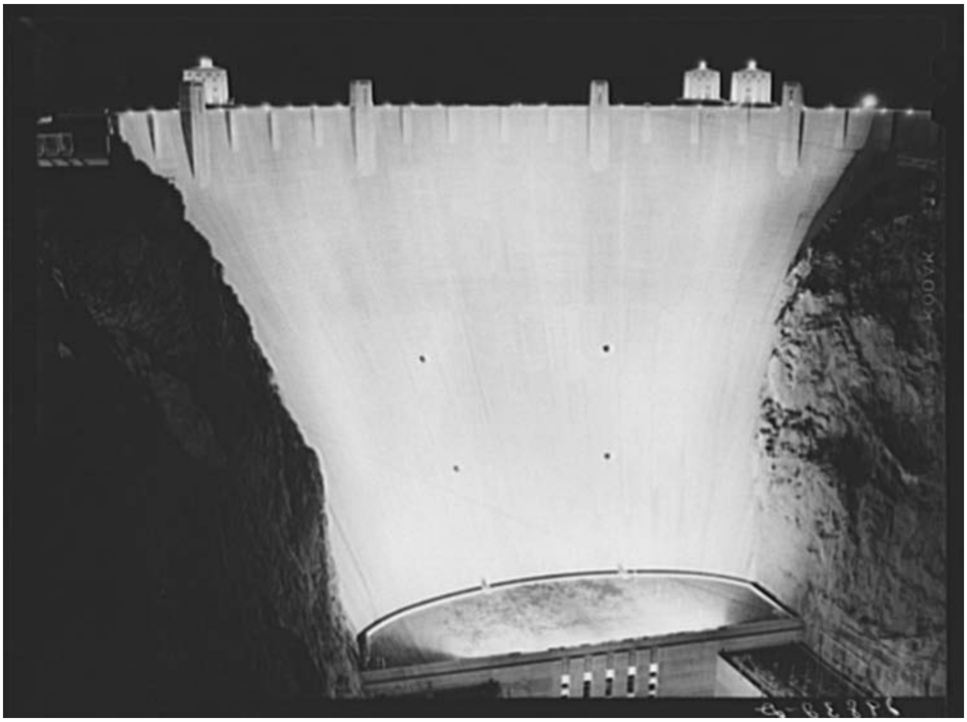


Figure 2. Hoover Dam.

the inland Chongdong region to commerce. Plans were first developed in the 1920s. They were revived in the mid-1950s, when devastating floods along the Yangtze raised the recurring need for new water policy measures. The project was finally started in 1993. The dam, almost 2.5 kilometers wide and 180 meters high, created a reservoir 66 kilometers long and hundreds of meters deep, at a cost of approximately $24 billion. On its completion in 2009, the first set of generators will produce an amount of power expected to create as much electricity as 18 nuclear power plants.

Controversy over Three Gorges Dam reflects the larger social, environmental, and economic questions of dam building. In the late twentieth century, the desire to protect rivers, riparian habitat, and human health hindered the construction of large dams and even mandated the removal of some. In China, critics of Three Gorges Dam claim that slowing the flow of the Yangtze River creates an environmental hazard by allowing sources of pollution to collect. Indeed, the inevitable buildup of silt behind large dams poses a technological problem that has not yet been solved by engineers. Cultural and social concerns also raise questions about the viability of large dams. The reservoir created by Three Gorges Dam drowned valuable antiquities sites, including evidence of the ancient Ba people. It also flooded more than 100 towns, forcing the resettlement of 1.13 million people—the largest resettlement in the history of dam building. Three Gorges Dam, like all major dam projects, reflects the need to balance industrial growth with environmental and social concerns.

See also **Concrete Shells; Irrigation Systems; Hydroelectric Power Generation**

JESSICA B. TEISCH

Further Reading

Downing, T. and McGuire G., Eds. *Irrigation's Impact on Society*. University of Arizona Press, Tucson, 1974.
Greer, C. *Water Management in the Yellow River Basin of China*. University of Texas Press, Austin, 1979.
Jackson, D.C., Ed. *Dams*. Ashgate, Brookfield, 1997.
Justin, J., William, C. and Hinds, J. *Engineering for Dams*, vol. 3. Wiley, New York, 1945.
Schnitter, N. J. *A History of Dams: The Useful Pyramids*. Balkema, Rotterdam, 1994.
Smith, N. A. F. *A History of Dams*. Davies, London, 1971.
Wittfogel, K. *Oriental Despotism: A Comparative Study of Total Power*. Yale University Press, New Haven, 1957.
Wolf, D.E., Wolf, D.D. and Lowitt, R. *Big Dams and Other Dreams: The Six Companies Story*. University of Oklahoma Press, Norman, 1996.
Worster, D. *Rivers of Empire: Water, Aridity, and the Growth of the American West*. Oxford University Press, New York, 1985.

Dentistry

In the early twentieth century a visit to the dentist was usually for a toothache, cured by extraction or replacement of missing teeth that had already been extracted. Preventive and restorative technologies existed but were not widespread, and materials were limited. Fluoride was known to have prophylactic qualities against caries, a vaccine was being developed for periodontal disease, and a zinc phosphate cement was used to seal fissures in the occlusal surface of molars. Diagnostic procedures included x-rays but x-rays were not generally applied in dentistry until about 1916. Anesthesia, in the form of nitrous oxide gas or ether had been in use in dentistry since 1863 (earlier than in surgery). In 1904, Alfred Einhorn introduced procaine (trade name Novocain), an injectable local anesthetic, to the practice of dentistry. It was superior to cocaine, which had been used previously, in that it did not cause tissue sloughing and was not addictive. The combination of anesthesia and electrical drills prepared the way for increased extraction and restorative work.

Public health leaders had an interest in dentistry, and throughout the world clinics were opened by philanthropists following the example of American industrialist George Eastman. By 1925 the inverse relationship between caries, or tooth decay, and tooth mottling as a result of natural fluoride in drinking water impelled further epidemiological studies, and by 1962 many U.S. cities and up to 30 other countries were adding sodium fluoride to their water supply. Fluoride treatments in the form of topical applications were provided for children, and dental hygienists visited schools showing students how to brush and floss their teeth. As a result of better dental education, the prevention of caries and dental restorations helped improve overall public health. Dental specialties such as orthodontics, endodontics, oral surgery, prosthodontics, pedodontics, and periodontics rapidly developed, each with their specialized technologies. Along with local anesthesia, topical numbing, use of a suction machine, and a reclining chair with leg rests, psychological amenities such as music or television available to distract and ease the patient's anxiety and ability to undergo lengthy procedures if necessary.

Technologies and Materials

Technologies used in diagnostic work include x-rays, study models, and an examination using explorers and probes to find cavities and measure periodontal pocket depth. The first x-ray equipment specially designed for dental radiography was introduced around 1923, but early machines were not shockproof. The modern dental unit in which x-ray tube and high-voltage transformer are both contained in a single comparatively small housing was introduced in 1933. Early x-ray films were made from celluloid but had to be cut and wrapped, and were dangerous because of their flammability. The x-ray machine evolved with film technology. As films became faster, exposure to radiation for both dental personnel and patient could decrease. A long tube replaced the short cone in order to increase the distance between target (patient) and film.

Periapical and bite-wing views require 16 films and show individual teeth. Oral surgeons, orthodontists, and prosthodontists require additional information about oral pathology, impacted teeth, developmental dentitions, jaw relationships, and the temporomandibular joint as part of their diagnostic work. In the 1950s, a panoramic machine was built that traveled around the patient's head. It only required one exposure to radiation. The image produced is on a large film and shows the complete dentition, each root and below, along with occlusal relationships. By the end of the twentieth century, digital intraoral photography that instantly magnifies the teeth and underlying structures on a computer screen had been invented.

Early formulas for impression materials contained shellac, talc, glycerin, and fatty acids. A superior product, hydrocolloid (reversible agar–agar material) was heated, poured into trays, then cooled in the mouth and used to make the initial impression for inlays and crowns. After the impression was made, it was filled with a type of dental stone more accurate than plaster of Paris for the model. Today, a study model impression is taken with alginate, an irreversible colloid material, mixed with water. This is placed in a tray in the patient's mouth, removed when set. Plaster of Paris, a gypsum hemihydrate compound, is poured into this shell to create the model.

Restoration of teeth, whether with an amalgam filling, inlay, or crown, requires preparation with the use of an abrasive dental drill. The clockwork motor-driven drill was first replaced by electric plug-in drills in 1908. These early drills operated at 600 to 800 revolutions per minute (rpm). By mid-century a high-speed hydraulic drill, which operated at 60,000 rpm, was developed by the National

Institute of Standards and Technology (NIST) and the American Dental Association. Modern water-cooled, turbine-powered drills rotate at speeds between 300,000 and 400,000 rpm. An assortment of different shapes of burrs that fit the head can be interchanged for ease of manipulating the shape of the preparation.

The first modern materials for restorations of small cavities were made in the early nineteenth century from gold foil, and from the 1900s, gold alloys. The process required special pliers, an alcohol lamp, a gold foil annealer, and a mallet. The process was lengthy and very uncomfortable for the patient. Silver–mercury amalgams were first used from 1832, but the poor quality limited use until a standard manufacture was developed in 1895. Modern amalgam is made from mercury mixed with silver, copper, zinc, and tin. As it hardens, the shape of the occlusal surface is carved and then burnished and polished after it assumes its final hardness. Baked porcelain inlays had been in use from 1862, once an effective dental cement had been developed. By mid-twentieth century, tooth-colored silicates or synthetic porcelain replaced gold and provided an alternative to amalgam. The powder and liquid was quickly mixed on a glass slab to a thick consistency with a spatula, and then transferred to the tooth. The material was smoothed and after setting, polished with a special attachment to the dental handpiece. Bonding, a process using composite resin materials, has replaced older techniques and materials. The revolution in laser technology has allowed for gentle application of a polymer/monomer that creates a tooth shade indistinguishable from the natural tooth that neither discolors nor decomposes.

When a large segment of tooth is destroyed, crowns and bridges are fabricated. Early crowns and bridges employed gold alloys joined to gold. In mid-century, crowns made of porcelain fused to gold were developed. This was esthetically superior to the gold crowns and stronger with respect to the chewing surface.

The filling of a tooth root canal is necessary when decay has invaded the internal tooth structure. Root canals were first filled with silver points, which were cut and measured to fit the walls of the excavated tooth. An instrument called a reamer was used to clean out the canal, followed by filing and irrigation to clean and shape the interior portion. Gutta percha, an inert material made from the latex of a tree from Malaysia, gradually replaced silver as the material of choice for root canal filling.

Permanent and Removable Restorations

A denture is a prosthesis, an artificial device, made to take the place of missing teeth. It is either partial or full, and the appliance is either permanently cemented in the mouth or a clasp is made to hook the partial restoration onto existing teeth. In 1930, Vitallium, a chrome–cobalt alloy was introduced for making these clasps. Replacements for decayed or lost teeth have been produced for millennia. Charles Goodyear's invention of vulcanized rubber, which could be molded against a model of a patient's mouth, allowed cheap and self-retaining denture bases from 1864 (earlier false teeth on gold bases had been held in place by springs). In 1919, gingival-colored base material helped to make a more realistic prosthesis. Although implants were being fabricated from pyrolytic carbon in 1919, they did not gain widespread application until the 1990s. Despite high success rates, they are still considered experimental. Dentures today (teeth and base) are made from an acrylic material (methylmethacrylate) that is colored to simulate gum tissue and the tooth. A veneer or laminate is made from porcelain to fit the facial surface of a tooth that has been discolored or chipped or is otherwise esthetically unacceptable.

Dentures were improved through the invention of the articulator (a dental machine that works as closely as possible to the way the mouth works) by the Swiss Alfred Cysi in 1909. Esthetics developed to match the patient's face with tooth shape.

Impressions for dentures were taken with rubber mercaptan-based materials that came in two tubes, one with a paste and the other with an accelerator made from lead peroxide and sulfur. These were mixed, placed in trays, applied to the arches, and removed after setting. Silicone, and later acrylic, compounds replaced rubber base impression material as well as hydrocolloid. Today, impressions for dentures are made in the same way as study models.

Fixed restorations require retention by cement, usually composed of zinc oxide–eugenol or zinc phosphate compounds. An improved cement in the 1970s added polyacrylic acid, which made it less irritating to the dentine and stronger. Glass ionomers (zinc oxide and polycarboxylate) had the advantage of better retention and were also used as a filling material for anterior teeth. A silicate cement (fluoroaluminosilicate glass and polycarboxylic acid) was introduced in the late 1970s in England. Cements made from calcium silicate and organic acid are also used to line the floor of a cavity before filling. Bone cements, used for attaching implants, or prostheses that require

incorporation with osseous structures are now made from methacrylate and apatite compounds.

Orthodontia

Edward Hartly Angle established a school of orthodontia in 1900 and a year later established the American Society of Orthodontists. Orthodontics is the specialty that moves teeth to their optimum place in the oral cavity for both esthetics and function. The first bands used in orthodontia, made from stainless steel, were applied to each tooth and attached to wires that were tightened at timely intervals to move the teeth. Because these took up space in the mouth, it was routine to remove teeth. In the 1990s, braces made from very thin nickel titanium could preclude the need for extractions. To overcome resistance to treatment with metal braces, a new clear sapphire bracket was designed. A totally invisible lingual brace was designed for amenable cases. Cements used in orthodontia must adhere closely enough not to trap bacteria but also be removable at the appropriate time. Acrylic adhesives have made this possible.

After the teeth have been moved to their desired positions and the braces removed, it is necessary for the bone to grow back into the areas that have been altered. Removable appliances made from methymethacrylate are then prescribed until the teeth are solidly held in the jaws. Headgear, an external apparatus that caused many teenagers embarrassment, has been replaced with the use of tiny magnets.

For more complex problems, orthognatic surgery is sometimes undertaken to correct severe jaw problems, after which banding is performed to complete the esthetic result.

The AIDS Epidemic

The most revolutionary change in technology resulted after the mysterious case of dentally transmitted acquired immune deficiency syndrome (AIDS) in 1987 in Florida. After exhaustive research by the Centers for Disease Control, the young woman's case was traced to her HIV-infected dentist. The overhaul in protective gear for dental professionals now includes masks, eye protectors, gloves, and polyethylene disposable covers for the head, lamp, and everything that comes into contact with the patient. A rubber dam once used only for root canal procedures can provide a barrier between the tooth and saliva. Instruments are either sterilized or disposable. Although the mouth is not a sterile field and cannot be kept sterile, mandatory laws are enforced regarding the maintenance of dental offices and disposal of fluids, tissues, and supplies.

LANA THOMPSON

Further Reading

Asbell, M.B. *Dentistry: A Historical Perspective*. Dorrence, Bryn Mawr, Pennsylvania, 1988.

Barton, R.E., Matteson, S.R. and Richardson R.E., Eds. *The Dental Assistant*, 6th edn. Lea & Febiger, Philadelphia, 1988.

Foley, G.P.H. *Foley's Footnotes: A Treasury of Dentistry*. Washington Square East, Wallingford, PA, 1972.

Glenner, R.A. *The Dental Office: A Pictorial History*. Pictorial Histories, Missoula, 1984.

Guerini, V. *A History of Dentistry, from the Most Ancient Times Until the End of the Eighteenth Century*. Milford House, New York, 1909. Reprinted by Liberac, Amsterdam, 1967.

Langland, O.E. and Sippy, F.H. *Textbook of Dental Radiography*, 2nd edn. Thomas, Springfield, 1984.

Lyons, A. and Petrucelli, R.J. *Medicine: An Illustrated History*. Abrams, 1987.

Ring, M.E. *Dentistry: An Illustrated History*. Abrams, New York, 1985.

Seide, L.J. *A Dynamic Approach to Restorative Dentistry*. W.B. Saunders, Philadelphia, 1980.

Skinner, E.W. *The Science of Dental Materials,* 6th edn. W.B. Saunders, Philadelphia, 1967.

Stecher, P.G. Ed. *New Dental Materials*. Noyes, Park Ridge, 1980.

Weinberger, B.W. *An Introduction to the History of Dentistry With Medical and Dental Chronology and Bibliographic Data*, 2 vols. Mosby, St Louis, 1948.

Detergents

Detergents are cleaning agents used to remove foreign matter in suspension from soiled surfaces including human skin, textiles, hard surfaces in the home and metals in engineering. Technically called surfactants, detergents form a surface layer between immiscible phases, facilitating wetting by reducing surface tension. Detergent molecules have two parts, one of which is water-soluble (lyophobic) and the other oil or fat-soluble (lyophilic). They are adsorbed onto surfaces where they remove dirt by suspending it in foam. Some also act as biocides, but no single product does these things equally well.

The oldest detergent, soap, has been known since antiquity. Made by boiling animal fats and vegetable oils with caustic alkalis, soap is sodium or potassium stearate:

$$CH_3(CH_2)_{16}\text{—}COO^-\ Na^+$$

Lyophilic————Lyophobic

Soap combines with calcium and magnesium ions to form an insoluble scum in 'hard' water. For this reason synthetic detergents, less affected by hardness in the water, have replaced soap for many purposes.

Synthetic detergents were first manufactured in Germany in 1917 where they replaced soap, releasing animal fats for food and other uses during World War I. These early products were propyl- and butyl-naphthalene sulfonates, but in the late 1920s and early 1930s sulfonated fatty alcohols such as sodium lauryl sulfonate were found to be better detergents.

Sodium lauryl sulphate

$$CH_3(CH_2)_{10}CH_2 — OSO_3^- \ Na^+$$

Lyophilic Lyophobic

In America, where petroleum products were available, alkylbenzene sulfonates were used, and by the end of World War II these had displaced the alcohol sulfates as general cleaning agents, though the latter are still used in shampoos and fine detergents. Today alkylbenzene sulfonates account for about half of all synthetic detergents produced in the U.S. and Western Europe. Until the 1960s the alkyl group commonly used was tetrapropylene ($C_{12}H_{24}$). These detergents were marketed in both powder and liquid form; they have very good detergent properties and produce copious lather.

Sodium tetrapropylenebenzene sulfonate

$$C_{12}H_{24}.C_6H_4—SO_3{}^- \ Na^+$$

Lyophilic ——————————Lyophobic

As the large lyophilic part of all these molecules is negatively charged, they are called anionic detergents, by far the most common type. But there are also cationic, nonionic, and amphoteric detergents. Cationic detergents are widely used as textile fabric softeners. Some are germicides or fungicides and are used as sanitizers in hospitals. The nonionic alcohol ethoxylates are used in liquid detergents for the laundry, but they are not produced as powders. The amphoterics are used as foam stabilizers.

The apparently simple process of cleaning soiled surfaces proceeds in several stages. First, in reducing the surface tension, detergents allow the aqueous bath to wet the surfaces thoroughly. In the case of textiles they enable the water to penetrate the fibers, and in dyeing processes this helps the dye to spread evenly. Next, detergent molecules form a layer of electrically charged ions at the interface between the water and the surface to be washed. When the soil is bound to the surface by an oily layer, the detergent breaks this up into individual droplets, forming a colloid that is dispersed in the foam. The soil then passes into the washing water, usually facilitated by mechanical action and high temperature. Each of these stages has been scientifically investigated, with the results used to invent new detergents or to modify old ones to improve efficiency.

The largest quantities of synthetic detergents consumed in the household are spray-dried powders used in laundering fabrics. They are expected to produce a variety of effects in addition to their detergent action. Compounds that remove hardness from water are added to enhance the detergent power together with anciliary compounds such as bleaches, perfumes, and fabric brighteners. Protein stains such as egg, milk, or blood are difficult to remove by detergents alone and must be broken down by proteolytic enzymes to make them soluble or at least permeable to water. Hence in "biological" washing powders, the detergent is fortified with such enzymes to make it more effective. Proteolytic enzymes were first tried in washing powders in Germany in the 1920s with only moderate success, but improved strains of faster-acting enzymes were widely added to powders in Europe by the late 1960s, and later in the U.S. lipases and amylases are now added to break down fats and starches. Unfortunately these enzymes have a toxic effect on some people.

Useful as they are, the tetrapropylenebenzene sulfonates have caused serious environmental problems. The foam they produce remains on the surface of wastewater in drains and sewers. It retards biological degradation in sewage treatment plants and is difficult to remove in water regeneration. In rivers they reduce the solubility of oxygen in the water, killing plant life and fish. After intensive research in the 1960s the problems were reduced when simpler alkyl groups more easily broken down by bacteria replaced tetraproplyene.

Hard surface cleaners include products containing detergents for domestic, institutional, and industrial use. Some have added germicides, fungicides, or insecticides. They differ in strength according to their purpose and the way they are designed to operate, but all rely on the use of a mechanical scrubbing action, either manual or by machines. Domestic hard surface cleaners are used to remove a variety of soils from dried and baked-on food deposits (carbohydrates, fatty acids, proteins, etc.) to petroleum greases and oils. They

are designed to act efficiently on many different surfaces and are used in homes and offices to clean mirrors and windows, sinks and baths, bathroom and kitchen fixtures and appliances, table and counter tops, floor coverings, and patios. Cleaners for these purposes should act without damage to the surface and without leaving stains. Most are liquid and contain mild abrasives and solvents as well as detergents; those intended for use where food is prepared must be nontoxic.

Although detergents are usually thought of as water-soluble, there are also oil-soluble substances that hold foreign matter in suspension. Called dispersing agents (dispersants) rather than detergents, these substances employ the same physical principles but are strongly lyophilic and are used widely in engineering operations. They are added to lubricating oils in automotive engines to prevent the formation of hard deposits on cylinder walls and to inhibit corrosion. They are also added to petrol to prevent the build-up of gummy residues in the carburetor and to dry-cleaning fluids to facilitate the removal of oily soils from fabrics.

Special detergents also have many uses in industry and engineering. In the oil industry for example, surfactants are sometimes used to promote oil flow in porous rocks, or to flush out oil left behind by a water flood. In this case a band of detergent is put down before the water to create a low surface tension and thus allow the oil-bearing rock to be scrubbed clean. The water is often made viscous by adding a polymer to prevent it breaking through the surfactant layer. Detergents are also widely used in industrial flotation processes to separate lighter particles from a mixture with heavier materials.

See also **Laundry Machines and Chemicals**

N. G. Coley

Further Reading

Davidsohn, A.S. and Midwidsky, B. *Synthetic Detergents,* 7th edn. Longman, Harlow, 1987.

Kirk-Othmer. *Encyclopaedia of Chemical Technology,* 4th edn. Wiley, 1997, vol. 23, pp. 478–541.

Lange, K. R., Ed. *Detergents and Cleaners*. Hanser, New York, 1994.

Schramm, L. L. *Surfactants: Fundamentals and Applications in the Petroleum Industry*. Cambridge University Press, Cambridge, 2000.

Spitz, L., Ed. *Soaps and Detergents*. American Oil Chemists Society, Champaign, 1996.

Tadros, T.F. *Surfactants*. Academic Press, London, 1984.

Schwuger, M.J. *Detergents in the Environment*. Vol. 65 of the Surfactant Science Series. Marcel Dekker, New York, 1997.

Lai, K.-Y. *Liquid Detergents*. Vol. 67 of the Surfactant Science Series. Marcel Dekker, New York, 1997.

Showell, M.S. *Powdered Detergents*. Vol. 71 of the Surfactant Science Series. Marcel Dekker, New York, 1997.

Diabetes Mellitus

The disorder called diabetes mellitus, translated from the Greek, means "flow of urine like honey." In the second century AD Aretaeus the Cappadocian (or more likely one of his assistants) made the diagnosis by dipping a finger in urine and tasting the sweetness. By 1889 Josef von Mering had shown that diabetes could be produced in animals by removing the pancreas. It was known that the pancreas secreted a substance into the duodenum that was essential for absorption of food. Additionally, it was postulated that the pancreas had a function related to glucose control. By 1916 the substance that the pancreas released into the blood stream and that was central to control of blood glucose levels was named insulin although it had never been isolated. Frederick Banting and Charles Best were the first to produce a pancreatic extract that lowered blood sugar in 1922, and they received a well-deserved Nobel Prize for their discovery. The first patient to be treated was a 14-year-old girl in Toronto. Prior to this, what we now call type I diabetes had a mortality rate of 100 percent within two years. Few would dispute that the discovery of insulin was one of the outstanding achievements of the twentieth century.

Insulin is a pancreatic hormone that facilitates the uptake and metabolism of glucose from the blood into muscle where it is essential for energy requirements. Without insulin, blood sugar levels rise, and fat has to be used to supply the body's energy requirements. The end products of fat metabolism are highly acidic, and when the blood becomes acidic, coma and death ensue.

The original pancreatic extracts were crude and short acting, and they contained many impurities. Insulin was crystallized in 1926, and in the 1930s it was combined with zinc, protamine, or globin in various proportions to achieve different lengths of action. Protein purification procedures were developed in the late 1960s to produce monocomponent insulins. Between 1951 and 1971 Frederick Sanger and Dorothy C. Hodgkin worked on determining the structure of insulin, and their work laid the foundation for the next major development. A report in 1978 of an anticipated shortage of animal insulin led to the development of synthesized

human insulin. The two ways of doing this were either by enzymatic conversion of porcine insulin or by recombinant DNA technology. Human genes are synthesized for the insulin molecules A or B chains then joined by recombinant methods with sequences coding for β-galactosidase. Plasmids containing the modified genes are inserted into a special strain of the bacteria *E. Coli*, which after fermentation produce the insulin that is cleaved from the bacterial proteins and purified. Both these techniques produced a peptide indistinguishable from pancreatic human insulin. By the late 1980s, completely synthetic insulins were available that were characterized by manipulation of the insulin molecule to modify its pharmacokinetics (speed of onset and duration of action). These products were termed insulin analogs (1992) and could equally well be described as "designer" insulins.

A series of studies in the 1980s showed that tight control reduced the microvascular complications of diabetes. Thus, physicians were at last able to achieve good 24-hour control of blood sugar levels by using different types of insulin and multiple injection techniques to mimic the normal pattern of insulin secretion according to the patient's eating, sleeping, and exercise habits.

The treatment of type II diabetes (which afflicts older, overweight people) was difficult because of insulin resistance, a key feature of this type of diabetes, which made injected insulin less satisfactory. In 1942, during a typhoid epidemic in occupied Europe, researchers in a trial for a new compound of the sulfonamides (antimicrobial agents introduced in the 1930s) noted severe hypoglycemia (low blood sugar) in the subjects. The sulfonylurea group of drugs, which stimulate the pancreas to produce more insulin, were the result of further research; examples include tolbutamide, glibenclamide, and gliclizide. The diguanides were rather more physiological; their mode of action was to increase glucose uptake, but they were not free of side effects. The thiazolidinediones, introduced in 1996 (rosiglitazone and pioglitazone), also reduce insulin resistance and increase the uptake of insulin from the blood into the cells. As with the diguanides, this is a more physiological approach. By 2000 they had not been in use long enough to be assessed.

The first successful pancreas transplant was performed in 1967 by Dr. Richard C. Lillehei at the University of Minnesota, and is now usually performed at the same time as a kidney transplant for Type I diabetes patients with severe complications. Islet cell transplants are still experimental, although pig islet cells have been transplanted into immunodeficient mice and dogs, with rejection prevented by immune-suppressing drugs.

At the end of the twentieth century, diabetes was the prime example of self-management of a serious disease. Patients inject themselves with insulin, using disposable "dial-a-dose" syringes. They determine the dose by measuring their own blood glucose levels. This can be done by using a spring-loaded lancet to prick the finger and produce a drop of blood that is applied to a strip impregnated with a reagent area containing glucose oxidase. The resulting color change is measured by a small reflectance meter. This gives a blood sugar reading in millimoles per liter within 15 seconds. When these systems were introduced in 1978, there was some skepticism as to whether patients could cope, but they have proved highly adaptable with even five-year-olds rapidly becoming competent in measuring their sugar levels and taking the appropriate dose of insulin. Later developments included digital devices that were even easier to use.

The decrease in the complications of diabetes during the century has been due to several therapeutic advances. Renal damage and hypertension (high blood pressure) have been reduced by the use of the angiotensin-converting enzyme (ACE) inhibitors, which are effective in protecting the kidney and controlling blood pressure and thereby further reducing the risk of heart disease. Until the 1980s renal dialysis was not considered an appropriate treatment for the kidney failure of diabetics, but, fortunately, that rigid and inflexible attitude no longer applies. Erectile failure is a problem for diabetics, but even the U.K. government accepted that diabetics were allowed to have sildenafil (Viagra), which enhances penile blood flow, on National Health Service prescriptions. Laser photocoagulation of diabetic retinopathy was introduced in 1984. As a result of retinal screening and laser treatment, blindness is now rare in diabetic clinics. Finally, the pioneering work of D. Hockaday and K.G. Alberti in 1972 pointed to a physiological approach to diabetic coma using frequent very low doses of insulin, and mortality was reduced drastically.

See also **Diagnostic Screening; Hormonal Therapies**

JOHN J. HAMBLIN

Further Reading

Bliss, M. *The Discovery of Insulin*. Paul Harris, Edinburgh, 1983.

Day, J. *Living with Diabetes Treated with Insulin*. Wiley, New York, 1998. Written for patients. Excellent insight into modern management.

Hakim N., Ed. *Pancreas and Islet Cell Transplantation*. Oxford University Press, Oxford, 2002.
Jovanoviv-Peterson, L. and Peterson, N. *A Touch of Diabetes*. Wiley, London, 1998. Written for patients. Excellent introduction for nonmedical readers.
Pickup, J. & Williams, G., Eds. *Textbook of Diabetes*. Blackwell, London, 1991.
Wass, J. and Shalet, S., Eds. *Oxford Textbook of Endocrinology and Diabetes*. Oxford University Press, Oxford, 2002.

Diagnostic Screening

The early twentieth century physician used sight, smell, palpation, auscultation, and taste to aid in diagnosis of disease. Microscopy, chemistry, and x-rays were in their infancy, thus it was experience and the senses, not technology, that dominated the physical examination and the practice of medicine. There were few commercial laboratories until after World War I. Doctors returning from the battlefield benefited from their experience with wartime government diagnostic labs and promoted the value of having such facilities.

The technologies of diagnostic screening (imaging and biochemical technologies) developed concomitantly with discoveries in human physiology and pathology. The more subcellular information about a particular disease, the more specific a screening tool could be developed. For example, the complement fixation phenomenon was described in 1901 by Jules Bordet and Octave Gengou, and was later the basis of a number of diagnostic serological tests. Such tests determine whether or not a sample of serum has antibodies to a particular bacteria or virus. In 1905 Erich Hoffmann and Fritz Schaudinn discovered that the bacterium *Spirochaeta pallida* (now known as *Treponema pallidum*) was the causative agent in syphilis. Synthesizing this information, August Paul von Wassermann, Albert Neisser, and Carl Bruck used the complement fixation test to find *Treponema pallidum* in human blood before microscopy had allowed for visualization of extremely small organisms.

Although mammography was first used in 1927 to visualize palpable breast tumors, it was not until the late 1950s that Jacob Gershon-Cohen and Robert Egan demonstrated that cancers that did not manifest visibly could be detected with mammography. This led to the hypothesis that early detection of breast cancer by screening in asymptomatic women would or could prevent the cancer from spreading or increasing in size. In 1967 the machine consisted of an x-ray spectrum and tube mounted on a three-legged stand and a French company, CGR, produced the Senographe. By 1980, it was possible to reduce exposure time, and improve resolution and accuracy because of the chemical makeup of new films. A motorized compression device was added in that decade as well as an add-on component to assist in a breast biopsy should the procedure be necessary. In 1990, rhodium was added to the x-ray tube for better penetration of tissue by the ray. The last decade of the twentieth century heralded digitization in technology, and digital technology was incorporated in mammography to image spot views of suspicious breast tissue. When screening identified an area that required investigation, this additional view was taken. The digital cassette allowed for processing time to be reduced. Although most of the technical challenges of mammography have been solved, the procedure is best utilized in patients over the age of 50 because their breast tissue is generally less glandular and dense than that of younger women. The technological challenge for the next century will be to discriminate between dense normal and opaque pathologic tissue.

The first pregnancy tests in the 1920s screened for the presence of chorionic hormone in a woman's urine. The urine was injected into a rabbit, and after three days the animal would be killed and its ovaries examined for changes caused by human chorionic gonadotrophin (HCG). If changes were present, the woman was pregnant. As information about inherited diseases was developed, genetic screening tests were performed prior to conceiving. Sickle cell anemia testing, and a test for Tay–Sachs disease were available by 1971. In 1985, Unipath introduced a home pregnancy test that took 30 minutes and used litmus paper embedded with antibodies to test urine for the presence of chorionic hormone. Screening for genetic defects such as Huntington's disease (Huntington's chorea) has been the work of Elena Cattaneo in Milan and Erick Wanker in Berlin whose work screens for chemicals that prevent agglutination or clumping of neurons, a common characteristic of Huntington's chorea.

Screening for cardiovascular disease is based on the knowledge of certain risk factors associated with hypertension and heart disease. If these risks can be identified, patients can modify their lifestyles to improve their chances of a healthy life. Screening tests consist of the identification of chemical predictors in the blood such as cholesterol and by physical recordings such as electrocardiograms (ECG) and imaging technologies including echocardiograms and Doppler studies. The fore-

runner of the ECG was developed in 1901 by Willem Einthoven, a Dutch physiologist who created a rudimentary galvanometer using magnetic poles and a silver-coated string. The acceptance and further development of the technology were encouraged by the English physician, Sir Thomas Lewis (c. 1911), who stated that an examination of the heart was incomplete if this new method was neglected.

In 1928, S.A. Ernstine and A.C. Levine used vacuum tubes instead of a string galvanometer for amplification of the electrocardiogram. Until this time, the ECG machine was a table. One of the first portable models was made by Hewlett Packard. It weighed 27 kilograms and was powered by a 6-volt automobile battery. The next improvement was the incorporation of chest leads and then a central terminal (Wilson's Central Terminal) which could be placed anywhere on the body and connected to the other leads. In 1938, placement of the leads, wiring, and positions was defined in the U.S. and U.K. so that readings could be standardized. A few years later, Emanuel Goldberger increased the voltage on the central terminal and added three more leads, bringing the total to twelve. An expensive technology that did not use skin electrodes was developed in 1963 by Gerhard M Baule and Richard McFee, but because of its prohibitive cost, it was not adopted. By the early 1990s three additional electrocardiogram leads were used by researchers in Detroit, Michigan to increase sensitivity in the detection of myocardial infarction.

Combined technologies of electrocardiography, and Doppler ultrasound have allowed for the simultaneous imaging and recording of heart waves and anatomic information. An echocardiogram displays a cross-section of the heart and its major vessels as it beats. The Doppler records the direction and velocity of blood flow through the vessels. Most recently, researchers at Johns Hopkins University in Baltimore, have developed a technology for the nonspecialist that uses computer software from the Microsoft Access relational database to screen for heart and lung disease. The physician feeds five areas of information (stethoscopic, physical exam, electrocardiogram tracings, demographic, and health information) into the program and compares it to known diagnoses stored in the database.

One of the most common uses of diagnostic screening is for diabetes. Before technologies for diabetes screening, physicians tasted urine to detect sugar. When a patient with symptoms of frequent urination, weight loss, and persistent thirst presented, the test was to add Benedict's solution to urine to detect the presence of glucose. It required two test tubes, a medicine dropper, a heat source, and the reagent. In the 1950s, Ames developed Clinitest, a product that only required one test tube and a tablet. By 1960, a paper strip (Labstix) had been developed which could be dipped in urine to test for the amount of glucose. Ten years later, blood glucose monitors were available for home use. Called a reflectance meter, it weighed 1 kilogram and was 17 cm high and 10 cm wide. In the 1980s the Glucocheck device added memory to its product. At the beginning of the twenty-first century, there are 30 different types of diabetes meters, most developed in the 1990s with computer technology.

While there are no screens available for many types of cancer, for example lung cancer, diagnostic screening for cervical cancer was developed in the 1960s and where screening has been introduced in developed countries, mortality has been significantly reduced. The Pap smear (named for George Papanicolaou) is a simple test by microscope to examine cells for changes that signify disease. Screening for prostrate cancer remains poor, with cancer most often found by digital rectal examination of men who have no symptoms of the disease. A prostrate-specific antigen (PSA) test can detect a protein released into serum as prostate cancer grows, but only one third of men with prostrate cancer are positive.

See also **Diabetes Mellitus; Electrocardiogram; Genetic Screening and Testing; Histology; Ultrasonography; X-rays in Diagnostic Medicine**

LANA THOMPSON

Further Reading

Fye W.B. A history of the origin, evolution, and impact of electrocardiography. *Am. J. Cardiol.* 73, 937–949, 1994.

Reiser, S.J. *Medicine and the Reign of Technology*. Cambridge University Press, Cambridge, 1978.

Cartwright, L. *Screening the Body: Tracing Medicine's Visual Culture*. University of Minnesota Press, Minneapolis, 1995.

Russell, L. *Educated Guesses: Making Policy about Medical Screening Tests*. University of California Press, Berkeley, 1994.

Dialysis

Dialysis, or hemodialysis, is a process of separating substances from the blood. Three quarters of a million people worldwide suffering irreversible kidney failure are maintained today using what is commonly called the "artificial kidney," or hemodialyzer, the first machine to substitute for a failing

organ. Around 1960 and by coincidence two separate methods of treating irreversible renal failure developed almost simultaneously as practical treatments: dialysis and renal transplantation (see Organ transplantation).

The idea that death in uremia following kidney failure results from retention of solutes normally excreted by the kidneys emerged during the nineteenth century. Diffusion in gases or liquids was worked out in parallel, principally by Thomas Graham, with mathematical development by J. H. vant'Hoff. These two ideas underlie the possibility of treating kidney failure by dialysis of potentially toxic solutes from the blood, and hence the tissues and body water.

The first attempt to carry out *in vivo*, extracorporeal dialysis of blood was made in 1913 by John Jacob Abel (1857–1938) in Baltimore, Maryland, at the Johns Hopkins Hospital, together with Canadian physician Lawrence Rowntree and English biochemist Benjamin Turner. Extracorporeal circuits had been developed in the 1880s for perfusion of whole organs *in situ* or *ex vivo*. Dialysis needed a semipermeable membrane that would retain protein and cells but allow solutes to diffuse from the blood. A number of substances had been tried in the laboratory, but the best seemed to be collodion, developed for photography in the 1850s as sheets of cellulose dinitrate. Blood was difficult to dialyze since it clotted on contact with foreign surfaces, but about the turn of the century hirudin, the anticoagulant in leech saliva, became available in crude form, and Abel and colleagues used collodion and hirudin. Their design consisted of a number of collodion tubes immersed in a cylindrical glass bath of dialysis fluid; blood drawn from an artery was pumped through rubber tubes into the apparatus and back into a vein. They dialyzed dogs and showed urea could be removed, but they were more interested in separating physiologically active substances from blood. At a demonstration of their dialyzer in London in 1913, a staff writer at the London *Times* first used the term "artificial kidney." The team broke up after 1915 and Abel did no more work on dialysis.

The first specific attempts to treat uremia by dialysis were by Georg Haas (1886–1971) of Giessen, Germany, initially in ignorance of Abel's work. He built a machine very like Abel's in 1923 and treated the first human patients by dialyzing their blood *ex vivo* and reinfusing it intermittently. He also used hirudin as an anticoagulant, but this substance was still crude and often toxic. His patients had irreversible chronic kidney failure and all died. Haas abandoned the work until 1927 when the new anticoagulant heparin was discovered by Henry Howell, again at Johns Hopkins, together with medical student Jay Maclean and later, retired pediatrician Emmett Holt. But Haas' patients still died, and in the face of opposition from the medical establishment, he gave up for good in 1928.

In 1923 Heinrich Necheles (1897–1979) of Hamburg was dialyzing dogs deliberately made uremic, but with a machine incorporating a new design: a flat sandwich with multiple layers of another dialysis membrane, the peritoneum from calves. However, despite his interest in uremia, he used this technique only to prepare physiological extracts from blood in dogs.

Despite the use of heparin, dialysis stalled because collodion was too complicated to produce, too fragile, and too difficult to use. A new membrane, cellulose acetate, was introduced in 1908, marketed as "cellophane" in France in 1910. By 1930 the material was formed into tubes in Chicago by the Visking Company to make "skinless" sausages and was rapidly shown to be useful in the laboratory for dialysis. William Thalhimer (1884–1961) a New York pathologist and hematologist realized in 1937 from his experience of anticoagulating and storing blood that heparin together with cellophane tubing made a practical dialyzer a possibility. He built such a machine and dialyzed dogs, but as he did no clinical work himself he encouraged others to try it in humans. In the early 1940s four individuals tried and succeeded, each without any contact with others.

In 1944 Jonathon Rhoads (1907–2002) of Philadelphia built and used a dialyzer with cellophane tubing wound in spiral on a frame within a bath of dialysis fluid and used it on a single patient. The dialysis worked, but the patient died, and he did no more. Toronto surgeon Gordon Murray (1894–1976), an expert on heparin and a colleague of Thalhimer, worked on dogs from 1940 to 1946 to perfect a similar design. In 1946 his first patient was treated three times and survived her acute kidney failure. He did only a few dialyses over the next five years, also designing and using a flat-plate sandwich type of dialyzer. Nils Alwall (1906–1986) in Lund, Sweden, worked on rabbits for four years or more before achieving success in humans in 1946. His design was similar to Murray's but was notable in that it had an outer casing permitting the pressure around the spiral to be controlled so that ultrafiltration of excess fluid could be limited. Alwall continued working in dialysis until retirement and trained many people to use his first

machines, mainly in Europe but also in Australia, Israel, and even the U.S.

The best known of these researchers is Willem Kolff (1911–). In September 1945 he performed the first successful dialysis with recovery of the patient in Kampen, Netherlands. Beginning in 1943 he dialyzed 16 other patients, all of whom died. Many of them, however, had irreversible kidney failure, and some improved briefly. Although Kolff never did any animal work, for the first time he established the size parameters necessary for human use. As a result, his machine had a much larger surface area. He employed a spiral of tubing wound around a large horizontal drum, laying halfway in an open bath of dialysate, which rotated with a coupling from a Ford car engine to allow blood to flow into the cellophane tubing. Kolff believed passionately in dialysis, and during the late 1940s he built and gave away a number of machines, sent plans everywhere so people could build their own, and toured the world to advertise the new technique.

Initially facing worldwide skepticism, he went to Cleveland, Ohio in 1950. His machine was improved in Boston by physician John Merrill, surgeon Carl Walter, and engineer William Olson. Despite its many obvious disadvantages (e.g., large priming volume and uncontrolled ultrafiltration), it continued in use until the early 1960s. One major factor in acceptance of dialysis treatment during the 1950s was that, unlike the milder forms of acute renal failure where conservative management was possible, dialysis made a major difference to survival in injured soldiers with severe but temporary renal failure during the Korean War (1950–1952). New dialyzers that were much easier to use were introduced. These included the twin-coil kidney designed and built by Kolff and Bruno Watshinger of Vienna in only a couple of months in Cleveland in 1956, and the flat-plate sandwich design invented first by the chemists Leonard Skeggs (1918–) (who later invented automated clinical chemistry) and Jack Leonards, and, independently, Arthur McNeil in Buffalo, New York.

By the end of the 1950s physicians everywhere, having inadvertently started dialysis in patients with chronic irreversible disease, found a slow, miserable "second death" inevitable as the blood vessels vital for access became unusable. A solution came from the polymer industry in the form of plastic polyvinylchloride (PVC) electrical insulation tubing, which was already used for connecting patients to dialysis. Although it was first discovered in 1937, PTFE or teflon, which was not wettable and did not stimulate thrombosis in blood, became available in the 1950s. Seattle, surgeon Warren Wintershide pointed this out to the renal physician there, Belding Scribner (1925–). He then got an engineering colleague Wayne Quinton to bend the tubes using heat and make a shunt joining an artery and vein externally so that it could be closed off from the machine between dialyses but allow blood to flow through it continuously: Suddenly, long term dialysis for irreversible disease was possible. One or two patients who started on dialysis only a few years later in 1966 remained alive at the turn of the twenty-first century. The shunts were improved by the use of another new material, silicone rubber. Then in 1965, from experience with venipuncture while working in a blood bank, James Cimino (1927–) and his colleagues in New York introduced access to veins enlarged by a surgically created arteriovenous fistula. By the end of the decade, these had become almost universal, and external shunts were already obsolete.

Initially, the reusable and then (from 1968) disposable flat-plate and coil dialyzers already in use for acute renal failure were also used for long-term dialysis. However in the early 1960s engineers at the Dow company in Michigan learned to make hollow fibers less than 100 micrometers in diameter from numerous different polymers for proposed uses in water purification, and in medicine as membrane oxygenators. Dick Stewart (1917–), a physician and chemist working with Dow, suggested their use for hemodialysis, and the first capillary hollow-fiber dialyzers were used in humans in 1967. By 1980 they were the predominant type of dialyzer in use, and from 1990 onward, used almost universally. Their great advantage, apart from efficiency, is their small size: a 30 by 15 centimeter hollow-fiber dialyzer has the same capacity as Kolff's original rotating drum dialyzer, which was more than one meter long and half a meter in diameter.

Commercial cellulose acetate dialyzing membranes continued in use until the 1990s when replaced by more permeable and more biocompatible synthetic compounds such as polysulfone. In addition, these more permeable membranes (of which the polyacrylonitrile AN-69 of 1969 is the prototype) are increasingly used to exploit convective rather than diffusive removal of solutes, as pioneered by Lee Henderson (1931–) in the U.S. and Eduard Quellhorst in Germany in the 1970s. This technology has been particularly effective in treatment of reversible acute renal failure in very ill patients, using continuous hemofiltration throughout the day.

The other major change in dialysis machinery was the introduction of continuous blending of dialysate from concentrated salt solutions and reasonably pure (but not sterile) water, pioneered by Arthur Babb for Scribner in Seattle in the early 1960s, and now universal. In addition, various online safety measures, starting in the 1950s with a bubble trap to prevent air entering the circulation and the measurement of dialysate temperature and ionic strength, have continued to be introduced through the years.

See also **Intensive Care and Life Support**

J. STEWART CAMERON

Further Reading

Cameron, J.S. *A History of the Treatment of Renal Failure by Dialysis*. Oxford University Press, Oxford, 2002.

McBride, P.T. *The Genesis of the Artificial Kidney,* 2nd edn. Baxter Laboratories, Chicago, 1987.

Dirigibles

From the time of the first balloon ascents in France in 1783, the problem of steering airships became obvious because of dependence on the wind for direction and on gas quantity and pressure inside the envelope for altitude control. It would take over a century of experimentation with gas inflation, propulsion, and altitude and directional control for airships to become practical.

The wildest solutions were proposed to solve the problem of dirigibility, but serious, scientific-based proposals also appeared. One of the first was that of Jean-Baptiste Meusnier's 1783 ellipsoid-shaped balloon, which was never built. His design included the concept of a ballonet, a gas container within the envelope (with the space in between filled with air) to keep the hydrogen and of a mechanical valve atop the balloon intended to release hydrogen on command, thus allowing better control of ascent and descent phases. While both concepts would eventually be incorporated into both balloons and dirigibles, the matter of propulsion remained difficult to solve.

The first steerable airship was flown in September 1852 by Henri Giffard but, like machines other pioneers flew over the next half century, it was handicapped by the heavy weight of the engine and accompanying fuel, to the point where there was scarcely enough space on board for a crew. With the invention of the first gasoline engine in 1896, the situation began to change, and several projects took shape by 1900, thereby defining three main types of dirigible: the nonrigid, the semirigid, and the rigid.

The nonrigid machines were the first ones designed in the nineteenth century and the most commonly in use. The envelope's shape was maintained through internal gas pressure (with the hydrogen itself placed within ballonets inside the hull). The crew section and engine(s) were attached by ropes to the envelope. Carrying capacity was limited, however, and there was a risk of hull buckling if pressure dropped, a fairly common incident. Cheap to build, such models were the most popular among inventors and armed forces eager to develop new observation platforms to supplement ballooning units.

The semirigid system attempted to solve the buckling problem by attaching the passenger and engine nacelles to a bottom keel on top of which the hull was stuck and inflated. Capable of carrying heavier loads than a nonrigid type, a semirigid airship could still be deflated and carried about the field. This system formed the basis for the "blimps" used for advertising through the twentieth century.

The rigid airship solution consisted of either a solid hull (as in the case of a couple of aluminum-clad airships) or an internal structure of girders made out of wood or aluminum. Due to the substantial size required to balance the weight of the airship with the volume of hydrogen needed, the rigid airship could carry the heaviest loads at the greatest speeds over the longest distances without buckling. The main proponent of this formula was a German nobleman, Count Ferdinand von Zeppelin (1838–1917), who argued for an armament program that would include rigid airships beginning in the late nineteenth century. In 1900, he flew his first machine, LZ-1, and through a process of trial and error, he and his engineers improved the design. By July 1908, the fourth zeppelin (as rigid airships soon became known) had successfully accomplished a 12-hour flight with 11 people on board, at a time when airplanes carried at best two people for half an hour. This and the popularity of Count Zeppelin in Germany helps explain why the rigid airship seemed assured a solid place in the development of aviation.

When World War I broke out, several nations built dirigibles for maritime patrol and artillery assistance, but rigid airships soon became the first strategic bombers, as they attacked Paris, then left the French capital to artillery and used planes to focus on London. Though failing in their intended use as destroyers of industrial infrastructure, these machines had a tremendous psychological impact on the civilian population, and in the long run

convinced advocates of strategic bombing of the advantages of some form of terror warfare. Yet the rigid airship (whether zeppelin or Schütte–Lanz, the latter made of wood) had also proven far too weak in the face of weather, limited speed, and the ever-improving performance of the airplane.

Although Germany was initially banned from building even transport airships after the war, an exception was made for the LZ-126, a model delivered to the U.S. Navy at Lakehurst, New Jersey, in 1924. In the meantime, The Royal Navy's R-34, a model extrapolated from the zeppelin formula, successfully flew across the Atlantic in 1919. These events forecast the advent of the dirigible as a long-distance transport machine. Its golden age came with the first flight of the LZ-127 *Graf Zeppelin* in 1928 and ended with the fiery crash of the LZ-129 *Hindenburg* in 1937. During that time, these airships contributed to the establishment of links across the North and South Atlantic and made possible comfortable passenger transportation at a time when long-range airplanes were useful for mail, but ill-suited for an elite clientele. With the outbreak of World War II, advocates of the airship, who had relied on a combination of technical arguments and public support, had to give in, whether in Germany or in other countries.

Other nations, too, sought to take advantage of the dirigible. Great Britain planned the opening of a link to India, but the airship chosen, R-101, was rushed into service and crashed on her inaugural colonial flight in 1930. The U.S., too, faced disaster, as it lost three of its airships in storms in the 1920s and 1930s. The U.S. Navy turned instead to using blimps for submarine spotting and convoy escort and kept them in service until the early 1960s.

At the beginning of the twenty-first century, the airship experienced a rebirth of sorts. After being limited to advertising endeavors for many years, the discovery of new construction materials and the development of new missions (such as permanent coastal surveillance) have prompted the development of new projects. The Zeppelin company, which became an industrial concern after World War II, sponsored the development of a small passenger machine for pleasure cruises, while the Cargolifter corporation began to design of a new kind of sky crane intended to carry very heavy loads. While the success of such endeavors is hard to gauge, its popularity in media reports reflects the public fascination with great technology projects, as dirigibles once were.

GUILLAUME DE SYON

Further Reading

Dick, H.G., and Robinson, D. *The Golden Age of the Great Passenger Airships, Graf Zeppelin and Hindenburg*. Smithsonian, Washington D.C., 1985.

Higham, R. *The British Rigid Airship 1908–1931: A Study in Weapons Policy*. G.T. Foulis, London, 1961.

Khoury, G.A. and Gillett, J.D., Eds. *Airship Technology*. Cambridge University Press, Cambridge, 1999.

Meyer, H.C. *Airship Men, Businessmen, and Politics, 1890–1940*. Smithsonian, Washington D.C., 1991.

Robinson, D.H. *The Zeppelin in Combat: A History of the German Naval Airship Division, 1912–1918*. G.T. Foulis, London, 1971.

Robinson, D.H. *Giants in the Sky: A History of the Rigid Airship*. University of Washington Press, Seattle, 1973.

Smith, R.K. *The Airships Akron and Macon*. U.S. Naval Institute Press, Annapolis, 1965.

de Syon, G. *Zeppelin! Germany and the Airship, 1900–1939*. Johns Hopkins University Press, Baltimore, 2002.

Topping, D. and Brothers, E., Ed. *When Giants Roamed the Sky: Karl Arnstein and the Rise of Airships from Zeppelin to Goodyear*. University of Akron Press, Akron, 2000.

Dishwashers

A dishwasher is an appliance that automatically cleans, rinses and dries dishes and utensils. As is the case for many other major appliances, the basic principles and elementary technologies for the dishwasher existed in the nineteenth century. At the close of the twentieth century, household dishwashers were still considered a luxury item, owned by a minority of the population even in countries with full electrification, indoor plumbing, and near universal ownership of refrigerators, washing machines, and vacuum cleaners. Even the most sophisticated models at the end of the twentieth century could not remove accretions of baked-on food, thus ensuring the continuity of some form of hand washing. It remains common even in developed countries to wash, rinse, and dry dishes by hand.

The first dishwasher patents were filed by Americans Joel Houghton in 1850 and L.A. Alexander in 1865 for machines intended for commercial use. Houghton's device was a wooden tub that used a hand-turned wheel to splash water over the dishes, ultimately with little cleaning effect. Alexander, improving upon earlier collaborative efforts, patented a tub mounted on feet at the bottom of which was a set of blades that would be rotated by a hand crank located outside the tub. The blades would splash water against and between dishes set tangentially above them in stationary wire racks. Both machines still required a considerable amount of human labor. They needed to be filled with and drained of water and

the crank turned by hand, arguably more work than the task required without a machine.

In the 1880s, Josephine Garvis Cochran of Shelbyville, Illinois designed a hand-operated mechanical dishwasher for use in her own domestic environment. The design consisted of a wooden tub containing a wire basket to hold dishes. The dishes would be rotated during the cleaning process. Hot water was sprayed into the tub and driven over dishes by plungers when a handle attached to the tub was turned. This model was later fitted with a motor. A motor was also added to the first commercially available hand-cranked model, displayed in New York in 1910 by the Walker Company. In Britain, similar models called the Polliwashup and the Blick were filled with water from a boiling kettle. The latter could be turned into a table by lowering its lid.

From the 1850s through the close of the twentieth century, dishwashers have followed the principle of propelling hot soapy water against dishes to clean them. The gradual introduction of plumbing and electricity into middle- and lower-class homes and the automation and mass production of the dishwasher by the 1950s made it available and potentially desirable to many consumers. The first freestanding dishwasher with permanent plumbing was introduced in 1920. Portable dishwashers were introduced in the 1930s that could be hooked up to the kitchen faucet or tap. Accessories such as the "aerated faucet-flo" (General Electric's term) provided an alternative faucet so that owners could access water without stopping and uncoupling a running dishwasher in midcycle.

By the middle of the twentieth century, fully electrified, large, industrial dishwashers were being marketed by companies like Toledo and Blakeslee. These galvanized iron or stainless steel machines could wash up to 11,400 dishes per hour, often using a conveyor belt. They had tank capacities of 160 liters for a wash, could pump up to 530 liters of water per minute, and weighed around 1000 kilograms.

During the latter decades of the twentieth century, manufacturers focused on refinement rather than invention. Dishwashers continued to spray hot water at high pressure onto dishes, most often by two rotating spray arms. Dishwasher interiors were made of either plastic or, in more expensive models, stainless steel. In the 1960s, exterior colors such as Avocado and Harvest Gold were introduced for dishwashers and other major kitchen appliances. By the 1960s manufacturers had introduced the built-in dishwasher, increasingly found in new homes. Built-ins came with a dwelling and stayed when it changed owners. They were situated near the sink so that the dishwasher drain hooked up to the sink drainpipe, and the entire machine connected to the household plumbing and electricity. With the advent of the built-in, front-loading dishwasher, the popularity of top-loading machines fell.

Late twentieth century affluent dishwasher owners could hide their dishwashers inside drawers paneled to match the kitchen cabinetry, with controls hidden inside. Electronic dishwashers offered timers that allowed washing to be delayed for up to nine hours and digital indicators to monitor the cycle's progress and to highlight problems. Noise reduction, introduced in the 1960s, reached its height in the 1990s and was achieved by wrapping the wash cabinet in layers of "insulation" consisting of natural and synthetic materials that would muffle the sound of the motor and water propellers.

With concern rising over the impact of dishwashers on the environment, manufacturers began to emphasize environmentally friendly features such as low water consumption, "economical" detergent dispensers, and the air-dry option. The trend toward air drying was particularly noteworthy as a reversal of priorities expressed in material culture and technology. Manufacturers in earlier decades sold heat-drying options by playing on housewives' fear of water spots caused by air drying. Fuzzy logic-based, or "smart" appliances emerged from Japan in the 1980s. A fuzzy logic chip in the dishwasher abetted the conservation of water and energy by estimating the level of turbidity (soil level) of the load and controlling the cycles accordingly.

See also **Detergents; Electric Motors, Small**

JENNIFER COUSINEAU

Further Reading

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.

Davidson, C. *A Woman's Work is Never Done: A History of Housework in the British Isles 1650–1950*. Chatto & Windus, London, 1982.

Forty, A. *Objects of Desire: Design and Society from Wedgwood to IBM*. Pantheon, New York, 1986.

Gideon, S. *Mechanization Takes Command: A Contribution to Anonymous History*. W.W. Norton, London, 1948.

Hardyment, C. *From Mangle to Microwave: The Mechanization of Household Work*. Polity Press, Oxford, 1988.

Lupton, E. and Miller, J.A. *The Bathroom, The Kitchen, and the Aesthetics of Waste: A Process of Elimination*. Princeton Architectural Press, New York, 1992.

Strasser, S. *Never Done: A History of American Housework*. Pantheon, New York, 1982.

Webb, P. and Suggitt, M. *Gadgets and Necessities: An Encyclopedia of Household Innovations*. ABC-CLIO, Santa Barbara, 2000.

Yarwood, D. *Five Hundred Years of Technology in the Home*. B.T. Batsford, London, 1983.

Useful Websites

Badami, V.V. and Chbat, N.W. *Home Appliances Get Smart*: http://www.hjtu.edu.cn/depart/xydzzxx/ec/spectrum/homeapp/app.html

Domestic Heating

Most domestic space heating systems used in the twentieth century in Britain, continental Europe and the U.S. were developments of technologies introduced much earlier. At the beginning of the twentieth century, many homes, especially in Britain, were still heated by solid fuel open fires set in fireplaces with chimneys. Closed solid fuel stoves of clay, brick or tile or metal stoves developed from the seventeenth century onward were also common methods of domestic space heating across continental northern Europe and North America, especially during the early parts of the twentieth century. Such coal- or wood-fueled stoves are typically four times more efficient than open fires. In addition to more efficient designs of metal stoves, new types of domestic room heater spread into use in the twentieth century as new fuels became available.

Gas heaters were invented in the 1850s; Pettit and Smith adapted Bunsen's gas burner for domestic heating in 1856. Electric heaters became practical after the development in the early twentieth century (patented in 1906) of a nickel–chrome heating element that did not oxidize when heated. This American invention led to Ferranti's parabolic reflector heaters in 1910 (in the U.K.) and Belling's electric heater in 1912, in which nichrome elements wound on fireclay strips would heat the strips to incandescence. Such units provided radiant heat, in which the heat travels in straight lines warming objects in its path. On the other hand, electric convector heaters, first introduced around 1910, provide warmth by convection through the movement of heated air. Fan-assisted electric convector heaters, first commercialized in the U.S. by Belling in 1936, only became popular after 1953 when a silent tangential fan heater was produced in Germany. The storage heater is another type of electric convector heater also invented at the beginning of the twentieth century. A storage heater comprises a heat-conductive casing containing ceramic heat-retaining blocks that are gradually warmed by electric elements, usually overnight. The stored heat is then released, sometimes with fan assistance, the next day. However storage heaters did not spread into use until after World War II. In Britain they were not much used until after 1962 when low-cost night electricity rates were introduced. Storage heaters provided a relative low-capital cost alternative to the central heating system.

The widespread use in Britain of individual room heaters into the late twentieth century was due to the much slower adoption of central heating than in most other developed countries. In 1970 only 30 percent of British homes had central heating, rising to 90 percent by the end of the century. In continental Europe and the U.S, where living in apartments and heating with closed stoves was more common than in Britain, domestic central heating spread much more rapidly.

Central heating is a system that heats a building via a fluid heated in central furnace or boiler and conveyed via pipes or ducts to all, or parts of, a building. There are many types of central heating system depending on the fuel used to heat the fluid—usually coal, oil, or gas (although wood fired and electric systems exist)—and the fluid itself, which may be steam, water, or air.

William Cook was the first to propose steam heating in 1745 in England. At first steam heating was mainly used to heat English mills and factories, but its development for domestic use took place mainly in America, notably by Stephen Gold who was granted a U.S. patent in 1854 for "improvement in warming houses by steam." Many different designs of steam heating system were in use in America by World War I. Steam, however, never really became popular for domestic central heating due to its relative complexity and problems such as noise and fear of explosions. Steam piped under the streets from a central boiler was used instead for "district heating" of groups of residential and commercial buildings in large cities, especially in the U.S. New York's famous steam district heating systems were introduced in the 1880s. Many district heating systems were operated as "combined heat and power" (CHP) systems in which high-pressure steam was used first to drive electricity generating turbines and then distributed at low pressure, or as condensed hot water, to heat buildings. Utilizing the heat that otherwise would have been wasted increases the overall fuel effi-

ciency from about 30% in a conventional power station to over 70% for a CHP system. District heating and CHP systems declined in the U.S. after the 1950s, and by the early twenty-first century, were most common in Russia and Scandinavian countries, especially Denmark.

Given the drawbacks of steam for "wet" domestic central heating systems, by the early twentieth century the preferred fluid became water. This is heated in a boiler and conveyed via pipes to radiators in individual rooms (really these should be called "convectors" as they heat a space mainly by convection rather than radiation). Warm air central heating systems were also popular in the U.S., Canada, and Germany. In such systems air is heated via a furnace and distributed through sheet metal ducts to room vents. Initially hot water and air central heating relied on the density difference between the hot and cold fluid for circulation, so-called "gravity systems." Subsequently electric pumps and fans were introduced to assist circulation, thus allowing smaller pipe work and more compact designs. Whereas early systems could only be controlled by adjusting the rate of fuel burning, the introduction of central heating controls, starting with Honeywell's clock thermostat in 1905, allowed more efficient and convenient utilization of heat. The fuel most commonly used by central heating furnaces and boilers also shifted from coal at the beginning of the twentieth century to oil and gas by the 1920s. By the 1930s most of the components of the modern domestic central heating system were in place, for example, the typical British thermostatically controlled, pumped radiator system heated by a gas boiler, which also provides hot water for bathing and so on (see Figure 3).

Following the oil price increases of the 1970s and growing concerns about the environmental impact of burning fossil fuels, attention has turned to the development and adoption of more fuel-efficient domestic heating systems that reduce the heat loss from buildings and utilize renewable energy. The first approach includes the installation of "condensing" central heating boilers in which the flue gases pass over a heat exchanger before exiting, thus increasing energy efficiency from approximately 65 percent to 90 percent. Another example of fuel-conserving heating technology is the heat pump, which extracts heat from a low temperature source, such as the air or the ground, and boosts it to a temperature high enough to heat a room. Most heat pumps work on the principle of the vapor compression cycle. The main components in such a heat pump are a compressor (usually driven by an electric motor), an expansion

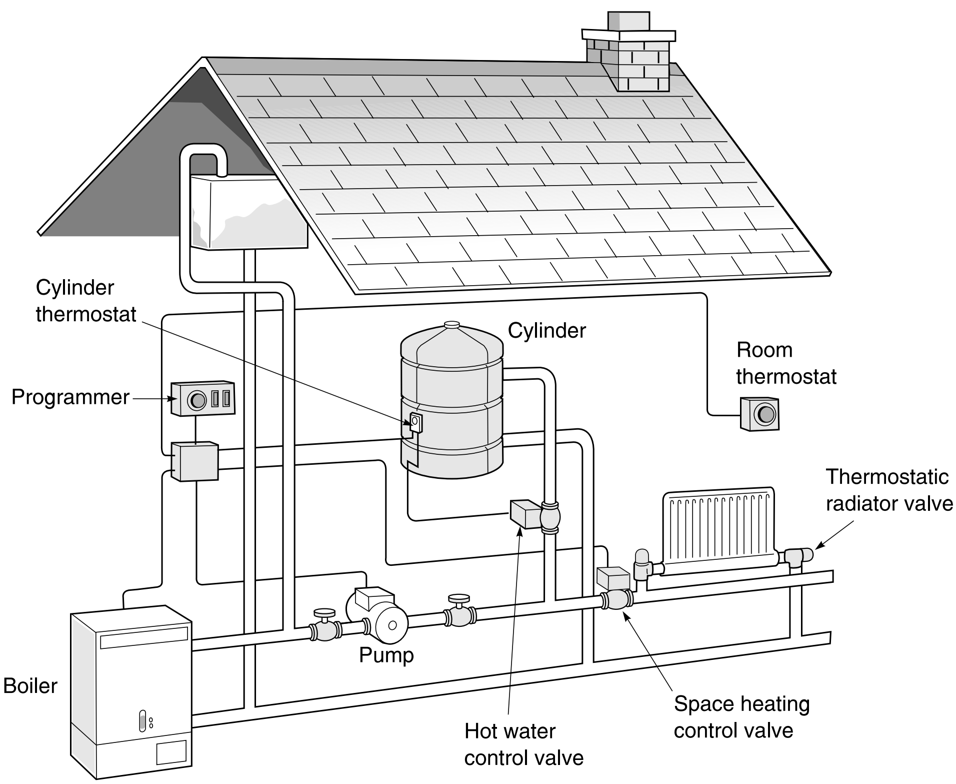


Figure 3. Typical "wet" or radiator central heating system, showing location of main components and controls.
[*From: Roy, R. with Everett, B. The National Home Energy Rating Activity, Supplementary Material to Open University course T172 Working with our Environment, Milton Keynes: The Open University, 2000. [Fig A6, p 60] © 2000 The Open University*.]

valve, and two heat exchangers, the evaporator and condenser, through which circulates a volatile working fluid. In the evaporator, heat is exchanged from the air or from pipes in the ground to the liquid working fluid, which evaporates. The compressor boosts the vapor to a higher pressure and temperature which then enters the condenser, where it condenses and gives off useful heat to the building. Finally, the working fluid is turned back into a liquid in the expansion valve and reenters the evaporator. Heat pumps were first demonstrated for domestic use in 1927 by Haldane in England and widely adopted in parts of the U.S. in the 1950s.

Solar energy for domestic heating, after experiments in the 1930s and 1940s, was given new impetus in the 1970s by the oil crisis and again in the 1990s by environmental concerns. Solar heating technology includes both "passive" and "active" systems. In passive systems, buildings are designed and oriented to maximize useful heat gain from the sun. In active solar thermal systems, the sun is used to heat a fluid (such as water) or air that can be used immediately or stored for space heating. Some solar homes provide space heating and hot water all year round without the need for any fossil fuel energy.

See also **Air Conditioning; Buildings, Designs for Energy Conservation; Solar Power Generation**

ROBIN ROY

Further Reading

Donaldson, B. and Nagengast, B. *Heat and Cold: Mastering the Great Indoors: A Selective History of Heating, Ventilation, Air Conditioning and Refrigeration from the Ancients to the 1930s*. American Society of Heating, Refrigeration and Air Conditioning Engineers, Atlanta, 1994.

Du Vall, N. Heating and housing, in *Domestic Technology: A Chronology of Developments*. G.K. Hall, Boston, 1988.

Harris, E. *Keeping Warm (The "Arts and Living" Series, Victoria and Albert Museum, London*, HMSO, London, 1982. Fully illustrated history of domestic heating equipment focusing more on visual and functional design than on technology.

Wright, L. *Home Fires Burning. The History of Domestic Heating and Cooking*. Routledge and Kegan Paul, London, 1964. A comprehensive, mainly nontechnical guide to the history of domestic heating and cooking from prehistory to around 1960.

Useful Websites

Nagengast, B. *An Early History Of Comfort Heating*. Other articles on the development of domestic heating systems may be found on this site: http://www.achrnews.com.

Dreadnoughts, *see* **Battleships**

Dyes

The decade commencing in 1900 marked the end of half a century of remarkable inventiveness in synthetic dyestuffs that had started with William Perkin's 1856 discovery of the aniline dye known as mauve. The products, derived mainly from coal tar hydrocarbons, included azo dyes, those containing the atomic grouping $-N=N-$, and artificial alizarin (1869–1870) and indigo (1897). By 1900, through intensive research and development, control of patents, and aggressive marketing, the industry was dominated by German manufacturers, such as BASF of Ludwigshafen, and Bayer, of Leverkusen. A new range of dyes based on anthraquinone (from which alizarin and congeners were made), and generally known as vat dyes, were the first major innovations in the twentieth century. Anthraquinone was obtained by oxidation of the three-ring aromatic hydrocarbon anthracene. Vat dyes are generally applied in reduced, soluble form; they then reoxidize to the original pigment and are extremely stable.

In 1901, René Bohn, head of the BASF Alizarin Laboratory, applied the indigo reaction conditions to a derivative of anthraquinone and discovered a blue colorant that he named indanthrone, from "indigo" and "anthraquinone." He then obtained the same product directly from 2-aminoanthraquinone. Later known as Indanthrene blue RS, it was the first of the anthraquinone vat dyes, noted for remarkable fastness.

Chemists at the rival Bayer company established the structure: indanthrone consists of two anthraquinone units joined through a ring of atoms containing two nitrogens. This enabled industrial research laboratories to discover anthraquinone-based intermediates for other vat dyes. In 1904, an assistant of Bohn synthesized benzanthrone, which afforded dibenzanthrone. Other important intermediates were anthrimides, readily converted into cyclic carbazole derivatives. The majority of anthraquinone derivatives were obtained by sulfonation of anthraquinone in the 1-, or alpha, position, and conversion to 1-aminoanthraquinone. The vat dyes, though expensive, in part because they required multistep syntheses, quickly became popular because of their resistance to light and washing, and were widely used in curtains, shirting fabrics, toweling, and beachwear.

It is worth mentioning that in 1915, stable complexes between azo dyes and a number of

metals, in particular chromium, were introduced for dyeing wool. Though generally duller than those of the unmetallized dyes, they were found ideal for suiting materials.

With the outbreak of World War I, the supply of German dyes to the major users in Britain and the U.S. ceased. This led to expansion of the then tiny U.S. synthetic dye industry, from which emerged that nation's organic chemicals industry. An important development in vat dye production was the building up of anthraquinone from naphthalene-derived phthalic anhydride. This was pioneered industrially in the U.S. during World War I, and removed dependence on the imported anthracene.

The absence of German dyes also spurred developments in Britain, where in 1915 Morton Sundour Fabrics of Carlisle produced the first anthraquinone vat dyes. From this endeavor emerged Scottish Dyes Ltd. of Grangemouth. In 1920, its chemists invented Caledon Jade Green, made from dihydroxydibenzanthrone (see Figure 4). It became the most widely produced vat dye in world. By 1928, when Scottish Dyes, through its acquisition by British Dyestuff Corporation, became part of ICI, almost 20 percent of dyes made in the U.K. were vat dyes.

The second major innovation in dye making took place at Scottish Dyes around 1932–1933. Chemist Arthur Dandridge and colleagues discovered a blue impurity when the intermediate phthalimide was prepared from phthalic anhydride and ammonia. The stable product was found to contain iron, and was studied at Imperial College, London, by Reginald P. Linstead and co-workers, who proposed a constitution similar to that of the pigment in chlorophyll. It was a phthalocyanine compound, of the type first prepared in 1907. ICI manufactured the copper analog, known as Monastral fast blue. Introduced in 1934, it represented the first member of the only new structural class of synthetic dyes in the twentieth century.

In 1925, the main German firms merged their interests to create the behemoth I.G. Farben. By this time the name reflected its historical roots rather than the range of activities. In the 1930s, the Bayer division developed a red azo dye into the first of the sulfonamide drugs.

The third important discovery in synthetic colorants satisfied the quest for dyes that attached to fibers by covalent bonds rather than weak intermolecular forces. This followed from research into wool dyes at the ICI General Dyestuffs Research Laboratory at Blackley, Manchester. William E. Stephen modified azo dyes by incorporation of reactive moities, particularly cyanuric chloride (trichlorotriazine). Failing with wool, Stephen suggested to Ian Durham Rattee in October 1953 that dyeing should be undertaken with cotton, which was successful. Thus was discovered the first fiber-reactive dye, of unprecedented fastness and introduced commercially as the

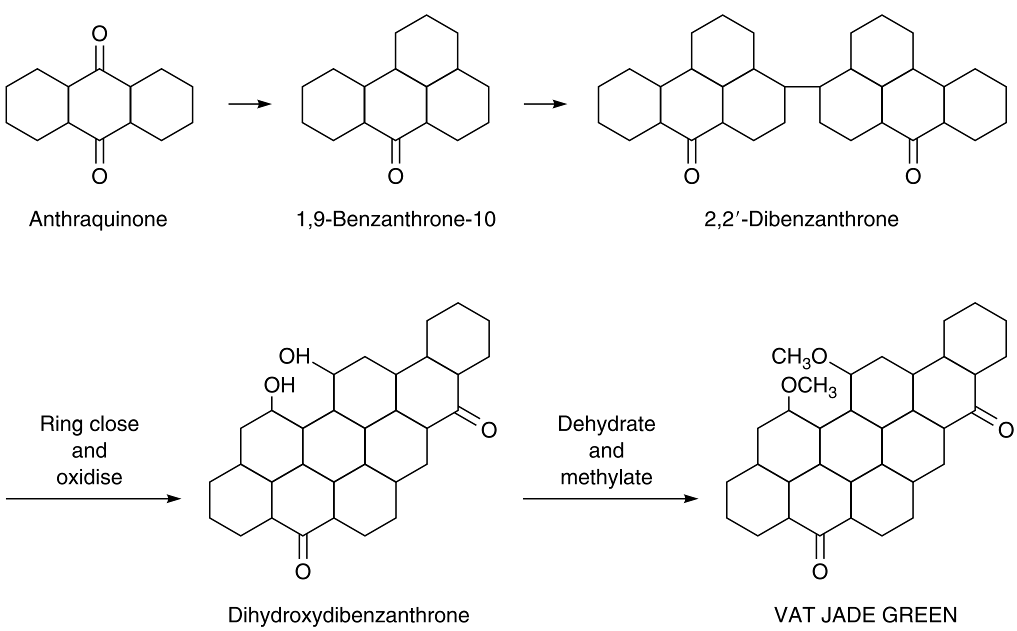


Figure 4. Flow chart for vat jade green.

ICI Procion range in 1956, on the hundredth anniversary of Perkin's discovery of mauve. The Swiss CIBA had already used the triazine grouping in dye synthesis, and the two firms came to an agreement over its application to reactive dyes. Fiber-reactive dyes displaced a number of vat dyes and reduced the incentive for research into new members of the latter class.

Semisynthetic and synthetic fibers introduced from the 1920s made new demands on the ingenuity of dye makers. Normally, dyeing is accomplished in aqueous solution, often in the presence of a fixing agent, or mordant. This is ideal for cotton, silk, and wool but not for certain synthetic fibers such as nylon and polyester that are plastic in nature; these require disperse dyes. The fiber is heated in an aqueous dispersion of a water-insoluble dye. Basic dyes are employed in the dyeing of polyacrylonitrile fibers.

While the two most important classes of dyes were anthraquinone vats and azo dyes, synthetic indigo became popular from the late 1960s with the swing towards fashionable denim and the faded look. Dyes are mainly used in textile printing and dyeing, but they have other uses, including food coloring and, when modified as pigments, for coloring plastics and synthetic fibers and for printing on paper. At the end of the twentieth century they found new and growing uses in the electronics industries, such as in inkjet printers.

From the 1970s dye making in the U.S. and Europe went into decline, due in part to reductions on tariff restrictions (U.S. and Britain) and environmental concerns (since many intermediate products were toxic). New centers for manufacturers were Asia, including Japan, India, China, and Eastern Europe.

See also **Coatings, Pigments and Paints; Fibers, Synthetic and Semi-Synthetic**

ANTHONY S. TRAVIS

Further Reading

Duff, D. A colourful tale. *Chemistry in Britain,* 37, 36–37, 2001. Celebrates a century of anthraquinonoid vat dyestuffs.

Forrester, S.D.A history of the Grangemouth dyestuff industry. *Chemistry in Britain*, 21, 1086–1088, 1985. The introduction of vat dyes into Britain.

Fox, M.R. *Dye-makers of Great Britain, 1856–1976: A History of Chemists, Companies, Products and Changes.* ICI, Manchester, 1987.

Haddock, N.H. Phthalocyanine colouring matters—their chemistry and uses. *J. Soc. Dyers Colourists*, 61, 68–73, 1945.

Lubs, H. A., Ed. *The Chemistry of Synthetic Dyes and Pigments*. Reinhold, New York, 1955. A comprehensive survey, with historical background.

Morris, P. J. T., and Travis, A. S. A history of the international dyestuff industry. *American Dyestuff Reporter*, 81, 59–100, 192–195, 1992.

Stead, C. V., Colour chemistry, in *Basic Organic Chemistry Part 5: Industrial Products*, Tedder, J. M., Nechvatal, A. and Jubb A. H. Wiley, London, 1975, pp. 315–50.

Steen, K. *Wartime Catalyst and Postwar Reaction: The Making of the United States Synthetic Organic Chemicals Industry, 1910–1930*. PhD thesis, University of Delaware, 1995. UMI Microfilm no. 9610494, 1996.

Travis, A.S. Synthetic dyestuffs: modern colours for the modern world, in *Milestones in 150 Years of the Chemical Industry*, Morris, P.J.T., Campbell, W.A. and Roberts, H.L., Eds. Royal Society of Chemistry, London, 1991, pp. 144–157.

Zollinger, H. *Color Chemistry*, VCH, Weinheim, 1987.

Zollinger, H. Some thoughts on innovative impulses in dyeing research and development. *Text. Chem. Colorist*, 23, 19–24, 1991.

E

Electric Motors

The main types of electric motors that drove twentieth century technology were developed toward the end of the nineteenth century, with direct current (DC) motors being introduced before alternating current (AC) ones. Most important initially was the "series" DC motor, used in electric trolleys and trains from the 1880s onward. The series motor exerts maximum torque on starting and then accelerates to its full running speed, the ideal characteristic for traction work. Where speed control independent of the load is required in such applications as crane and lift drives, the "shunt" DC motor is more suitable.

The electricity supply industry favored AC systems. DC motors could be adapted to operate on AC supplies, though not so well. The reversal of the current at each half cycle of the supply did not affect the direction in which the motor ran since both the field and the armature were reversed, but it induced "eddy" currents in the iron which wasted energy and heated the machine. Eddy currents could be limited by laminating all the iron, producing the "universal" motor that can be used on either AC or DC. This has been the motor used in most domestic electrical appliances, the exception being the washing machine, which requires greater power. The problem of eddy currents was also alleviated by working at a lower frequency. Some European railways employed AC series motors operating at 25 or 16 Hz rather the 50 Hz usually preferred for public electricity supply, but the railways were large enough undertakings to have their own generating stations.

One AC machine that could run from the public electricity supply was the induction motor. This was a new kind of machine, not a generator in reverse. Simple and robust, most of the world's electric motor power comes from induction motors. For most of the twentieth century, however, it had the serious limitation of being a fixed-speed machine.

DC Motors

Most DC motors have a fixed-field winding and a rotating armature with coils to which connections are made through brushes and a commutator. The windings are all on iron cores to enhance the magnetic field. As the armature and commutator rotate, the brushes make contact with successive commutator segments. The currents through those segments are switched as the commutator turns. In the series motor, the field and armature windings are in series, and the current flows through both. In the shunt motor, the windings are in parallel, and the current flowing through the field winding (which is of lighter construction than in the series motor) is controlled by an external regulator to adjust the speed of the machine.

The fact that DC motors and generators are essentially the same machines was recognized in the nineteenth century. Practical machines were introduced by Zénobe Gramme who, by 1874, had electrically driven machines in his Paris factory, although he used a single motor to turn a line shaft and not individual motors for each machine.

Large-scale use of electric motors began with electric railways and tramways in Germany, France, Britain and the U.S. in the 1880s. The deep "tube" underground railways and subways

were only possible with electric traction. The main pioneer of electric railways in Europe was the Siemens Company. In the U.S., Frank Julian Sprague (1857–1934) built a railway system in Richmond, Virginia, in 1888. Sprague's most important contribution was his development in 1895 of multiple-unit control, whereby separate motors distributed along a train were all controlled from a single operating position in the driver's cab.

London's first tube was the City and South London Railway, which ran initially from Stockwell to the City about 20 meters below the surface and opened to the public in December 1890. It had locomotives pulling separate carriages. Multiple-unit control with motors distributed along the train was introduced in London on what is now the Central Line in 1902.

The first mainline railway to adopt electric traction was the Baltimore and Ohio Railroad. In 1892 an extension to this steam railway included a long tunnel where steam traction would not be practical. Electric traction was used through the tunnel and was so satisfactory that contemporary observers thought it would be extended rapidly. In practice electric traction grew quite slowly.

There were different opinions about the ideal supply voltages for electric trains, and whether the supply should be AC or DC. In 1920 a Railway Advisory Committee appointed by the British government recommended the general adoption of 1500-volt DC, which remained the standard for some years.

The introduction of the mercury-arc rectifier in 1928 was a landmark in the development of railway electrification. It permitted AC supply to the train with rectifiers on the train feeding DC motors. Most British railways now use this system with 25 kilovolt overhead wires supplied from the National Grid. Since 1960, germanium and silicon power rectifiers have replaced the mercury-arc valves, resulting in much simpler, lighter, and more reliable rectifying equipment on the trains. The railways in southern England use third-rail DC at about 700 volts and have rectifying substations at intervals alongside the track. Other countries have used a variety of systems including AC at 16b and 25 Hz and also three-phase AC. The three-phase systems in Switzerland and Italy had one phase built into the track and two overhead wires for the other phases.

AC Motors

The universal motors mentioned above have been the usual choice of motor in small power applications such as sewing machines, food mixers, and vacuum cleaners. They can run at high speed, leading to a better power-to-weight ratio than other motors, and have been used in hand-held appliances such as electric drills since the 1920s. Almost every domestic motorized appliance except the washing machine uses a universal motor built into the appliance.

The first practical AC motors were the induction and synchronous motors developed by Nicola Tesla (1856–1943) in 1888; although the Italian, Galileo Ferraris (1847–1897), and others were working along similar lines. Tesla emigrated to the U.S. from eastern Europe, worked for a few years with Thomas Edison, and then joined the Westinghouse Company. Edison was a firm advocate of DC, and was firmly opposed to AC systems; George Westinghouse (1846–1914) was the leading American exponent of AC, and in changing employers, Tesla was stating his own views on the future direction of electrical engineering.

It had long been known that a pivoted permanent magnet or a pivoted piece of magnetic material would follow a rotating magnetic field. Tesla's great achievement was to produce a rotating magnetic field from alternating currents in two or more fixed coils. His first motor had two field coils energized by alternating currents whose waveforms were 90 degrees out of step. Tesla showed in 1888 that the resultant magnetic field was constant in strength but rotating in direction. He obtained American patents covering two-phase and three-phase induction and synchronous motors. (In the synchronous motor, the rotor is a permanent magnet and keeps in step with the rotating field. In the induction motor, the rotor lags slightly behind the rotating field, but the speed of rotation is still almost constant.)

Tesla also showed that an induction or synchronous motor could be run from a single-phase supply if part of the field winding were connected through a capacitor or inductor to produce a second phase. Once started, the motor would run satisfactorily on the single-phase supply; in many motors, the starting capacitor or inductor is switched out of the circuit automatically, usually by a centrifugally operated switch, once the motor is running.

The Westinghouse Company bought Tesla's patents, and from 1892 they began to promote polyphase AC distribution systems and the use of AC motors in industry. They adopted the three-phase 60-Hz electricity supply which remains the U.S. standard, although 50 Hz is preferred in Europe. For induction motors up to a few horse-

power, the simple and so-called "squirrel cage" rotor construction is normally used. For higher-rated motors, a wound rotor is generally preferred so that a resistance can be connected in series to reduce the starting current. The starting resistance is then cut out automatically when the motor has run up to speed. Induction motors are ideal where a constant speed is required. Because it needs no brush gear, the induction motor gives reliable service over long periods of time with little or no maintenance.

A motor that is easily confused with the induction motor is the repulsion motor. The confusion is compounded by the existence of mixed action motors that start as repulsion motors but run as induction motors. The repulsion motor is largely due to J. A. Fleming (1849–1945) and Elihu Thomson (1853–1937). Fleming, who was professor of electrical engineering at University College London and is best known for his work on the thermionic valve, made a study of the forces between conductors carrying alternating currents. In 1884 he showed that a copper ring suspended within a coil carrying an alternating current tends to twist so as to be edge on to the magnetic field. This is the basis of the repulsion motor.

Control of Motors

In many applications, motors are controlled simply by switching them on or off. If the starting current is heavy, then they may first be connected in series with a resistance, which is then cut out. Where several DC motors are working together, as in an electric train, it is usual to adopt series–parallel operation. If there are two motors and a resistance, then the controller will connect the machines to the supply in the following sequence:

1. Motors in series plus resistance
2. Motors in series only
3. Motors in parallel plus resistance
4. Motors in parallel

The principle can be extended if three or more motors are involved. When two motors are connected in series to the supply, each is effectively on half voltage, and the starting current is reduced accordingly.

In applications where precise control is needed over a wide range of speeds, then a DC motor and the Ward Leonard control system is widely used. H. Ward Leonard (1861–1915) was an American electrical engineer with a special interest in lifts. He appreciated that the ideal way of controlling a DC motor was to control its armature current. In the Ward Leonard system, developed about 1890, a first motor is used to drive a generator, and the output of that generator supplies the armature of the motor being controlled. The generator output is regulated by controlling its field current, which is of course very much smaller than the motor current. Numerous feedback systems have been devised in which the generator field current is controlled by a device that monitors any deviation of the motor's actual speed from its desired speed. Ward Leonard systems have often been used for winding motors in mines and for rolling-mill drives, but they have also been used in low-power applications where fine speed control is vital.

A two-speed variation of the induction motor is the pole amplitude modulated, or PAM, motor devised by Professor G.H. Rawcliffe in 1957. This is an induction or synchronous motor in which the field windings are so arranged that the effective number of poles can be changed by changing the connections of a few coils. The PAM motor therefore retains the reliability and robustness of the conventional induction motor while being able to work at either of two alternative speeds.

The modern approach to the ideal of an induction or synchronous motor whose speed can be varied is to provide a variable frequency supply, which was not feasible before the advent of power semiconductor devices. Transistors capable of controlling a few amperes became available in the mid-1950s, and by about 1960, thyristors (then known as silicon-controlled rectifiers) were available which could control some tens of amperes. Thyristors and other semiconductor devices are now available which can carry very large currents and switch to high voltages. The devices can also be connected together in series and parallel combinations so that there is no limit to the current and voltage they can control. They are used in inverter circuits to generate variable-frequency alternating current for supplying induction motors, and they are also used in "chopper" circuits, which switch a DC supply on and off rapidly to vary its effective voltage. Such control systems are easily regulated by reliable and compact electronic circuits that are responsive to speed or any other chosen parameter. At the end of the twentieth century, electric drives are commonly supplied as a unit incorporating motor and control equipment.

Electric Road Vehicles

Although only a small proportion of vehicles are driven by electric motors today, the electric vehicle

has an important history and is likely to become more important in the future because the electric motor offers a clean and quiet power source.

At the beginning of the twentieth century, electric road vehicles were more common than gasoline-driven ones. There were two distinct types: the self-contained machine supplied by batteries and the trolley vehicle that drew current from overhead wires. By 1900 there were several hundred electric taxicabs in New York and other U.S. cities and a smaller number in London. To encourage the use of electric cars, supplier fees for battery charging were kept low. The service provided by electric vehicles was considered quite adequate until World War I, when gasoline-powered vehicles were able to travel further without refueling. The first trolley bus service (a bus also powered through overhead wires but not limited to rails) began in 1901 in Bielethal in Germany, and trolley buses appeared in Britain in 1911. Just after World War II, the number of trolley buses in the world reached a peak of about 6000.

In some countries there are still many electric vehicles, usually small delivery vehicles, in daily use. Most electric vehicles use heavy lead–acid or nickel–iron batteries. Much research effort is currently going into the quest for a better battery. A possible contender is the sodium–sulfur battery, which offers a much lighter weight battery than conventional types. It suffers from the fundamental disadvantage that it operates at about 300°C, and hot sodium would be dangerous in an accident. Gasoline, of course, is also dangerous. It may well be that further research will find a battery, using the sodium-sulfur system or some other chemical combination, that will again give electric vehicles an advantage over gasoline-driven ones. Other research is looking at the fuel cell, producing electricity directly from either hydrogen or a hydrocarbon fuel but more efficiently and with less pollution. A further variant is the hybrid vehicle in which a small internal combustion engine drives a generator that charges a battery, and the battery supplies the motor driving the vehicle. The advantage of such an arrangement is that the internal combustion engine runs at constant speed and has only to supply the average power requirement of the vehicle, which is a small fraction of its peak power requirement.

Linear Motors

Linear motors are often described as ordinary rotary motors that have been split along their length and unrolled. It follows that there are as many kinds of linear motors as there are rotary ones. Linear motors may be AC or DC commutator machines, or they may be induction or synchronous machines, to name just a few possibilities.

The idea of a linear motor was described by several experimenters in the nineteenth century, but nothing was achieved before the twentieth. The Norwegian Kristian Birkeland obtained a series of patents between 1901 and 1903 for a DC linear motor used as a silent gun, and the idea of a linear motor gun has recurred periodically. The Russian engineer N. Japolsky worked on linear motors in Russia around 1930, and during World War II the Westinghouse Company in the U.S. built a linear motor aircraft launcher called the Electropult. The aircraft sat on a trolley with windings underneath, and the fixed part also had windings. The Electropult produced 10,000 horsepower and could accelerate a 4.5-ton aircraft to 180 kilometers per hour in 4.2 seconds.

The two potential applications of linear motors that have attracted most attention throughout the twentieth century are to drive shuttles in looms and in railways. Emile Bachelet set up a company to work on both applications in 1914. The main twentieth century figure on linear motors was Eric Laithwaite (1921–1997). He was first interested in the linear motor for use in a loom and turned later to its use in transport. The loom requires a means of projecting the shuttle at high speed across the width of the cloth. Conventionally, this is done by striking the shuttle very hard and catching it on the other side, a process that wastes energy and requires considerable mechanical strength in the shuttle. Despite many attempts by a number of inventors, linear motors have still not taken over in this application.

Laithwaite made considerable progress with linear motors for transport, especially with his concept of the "magnetic river," or "maglev" (for magnetic levitation), in which a magnetic field would lift, guide, and propel an object such as a train along a track without physical contact. A few small-scale linear motor driven transport systems are in operation at airports, but despite much research the large-scale linear motor driven mass-transport system remains a dream.

One application where linear motors have achieved success is in pumping and stirring liquid metals. In the 1930s pumped liquid metal was used for heat transfer in special circumstances, where the high thermal capacity of the metal was useful, and where direct contact with the liquid metal was to be avoided. The liquid metal itself acts as the

moving part of the motor, and all that is necessary to pump the metal is to fix a wound stator onto the wall of the container holding the metal. Such devices are now commonly used for stirring aluminum and other metals in furnaces and for assisting the transfer of metal to molds for casting, A specialized application is for pumping the liquid sodium coolant in nuclear reactors. Reliable operation and complete isolation of the metal being pumped are vital requirements, and the linear motor provides these.

See also **Electrical Power Distribution; Rail, Electric Locomotives; Urban Transportation**

BRIAN BOWERS

Further Reading

Hughes, A. *Electric Motors and Drives*, 2nd edn. Newnes/ Butterworth-Heinemann, Oxford, 1993.

Gottlieb, I. *Electric Motors and Control Techniques*, 2nd edn. McGraw-Hill/Tab Electronics, New York, 1994.

Gottlieb, I. *Practical Electric Motor Handbook*. Newnes/ Butterworth-Heinemann, Oxford: 1997.

Keljik, J.J. *Electric Motors and Motor Controls*. Delmar Learning, Albany, NY, 1995.

Valentine, R., Ed. *Motor Control Electronics Handbook*. McGraw-Hill Professional, 1998.

Yeadon, W.H. and Yeadon, A.W. *Handbook of Small Electric Motors*, McGraw-Hill Professional, New York, 2001.

Electrical Energy Generation and Supply, Large Scale

Public supply of electricity at the close of the nineteenth century was typically confined to the larger towns and cities where either a local entrepreneur, or a far-sighted municipality, established relatively small generating stations to supply local lighting loads. Many of these local power stations employed reciprocating engines to drive direct current (DC) dynamos. Overhead circuits generally carried the power no more than a kilometer or two to local businesses or the larger households in the district. Sometimes, where water-powered mills had existed previously, hydroelectric generators were established to supply the electricity consumers. As more and more people began to appreciate the convenience of electrical power and, moreover, could afford to pay for it, demand on local supplies increased and larger power stations began to be established. The invention of the electrical transformer to step-up the voltage at the generating station and step it down again to a safe level for use by the consumers, meant that higher speed alternators, often driven by steam turbines, could be employed to produce the power. High-voltage distribution reduced the losses in the circuits between the generating stations and the loads.

Development of the Grid

Large-scale electricity supply began in the second and third decades of the twentieth century with the establishment of so-called "grid" systems which provided for high-voltage interconnection of power stations. Such interconnection allowed spare generating plant on the system to be shared amongst a number of power stations. This avoided the previous need for individual power stations to carry a spare generating set to allow the consumer loads to be supplied when one of the other generating sets was out of service for maintenance or repair. The cost of providing spare capacity at each power station was becoming increasingly uneconomic as set sizes grew larger due to the strive for increased efficiency and economies of scale in their construction and operation.

Coal by Wire

By the middle of the century, the concept of a power station built in a town to supply only the load of that town had virtually disappeared, certainly in the U.K. At that time, more than 90 percent of electricity was generated in coal-fired power stations which were supplied either by coal trains or seagoing vessels. As high-voltage transmission technology developed at 220 kilovolts and above, it became more economical to locate the power stations on the coal fields themselves and transmit the electricity by high-capacity extra-high voltage circuits to the load centers. As the century progressed into the 1960s and the 1970s, individual power station sizes grew from the few hundreds of megawatts, to several thousand megawatts. Typical of the largest of the European coal-fired power stations is that at Drax in Yorkshire, which has a capacity of 4000 megawatts.

Generating Plant Technologies

Separate entries in this Encyclopedia describe the technologies of hydroelectric power generation and other renewable energy resources, turbines, nuclear reactors, and fossil fuel power stations. While coal was a dominant fuel in many countries for power generation over the first half of the century, some countries had other resources such as water, which is employed extensively in Norway and Switzerland. The great rivers of the U.S. and

Canada were also exploited to provide hydroelectric power. The generating plant at Niagara Falls operated throughout the century, but it was the 1930s and 1940s that saw the installation of many of the large hydroelectric dams such as Hoover (Boulder) on the Colorado River, Bonneville, and Grand Coulee on the Columbia River, Washington. Canada's large-scale hydroelectric plants, such as Churchill Falls in Labrador, were built somewhat later in the century. Hydroelectric power still provides for more than 50 percent of Canada's needs.

France in particular has developed nuclear power stations over the last 40 or so years of the century until they now supply over 90 percent of their electricity. The U.K. has developed many gas-fired power stations in the last decade of the century employing gas turbines, fuelled from North Sea gas, to drive the electrical generators.

Competition in the Industry

Most electricity supply systems in the developed world were originally state-owned, and many still are, except in the U.S. In the last few years of the century, however, notably in the U.K., there has been a move to break up the vast generation and distribution monopolies. The generation assets were divided up into several companies that would compete with one another on electricity price in a wholesale market, the cheapest generators being called upon to generate first. The grid system transmitted this wholesale power to distribution networks that supply individual consumers in a particular area. Both the transmission grid and the distribution network are owned by public companies with shareholders but are regulated as a monopoly service provider in terms of service and quality standards. Purchase of wholesale power and its resale to individual consumers, whether individual households or large industrial concerns, are undertaken by licensed supply companies that pay charges for the use of the transmission grid and the distribution networks.

Generating Efficiency and Combined Heat and Power

Because of limitations in the thermodynamic efficiency of steam turbines and gas turbines, the majority of the energy in the fuel supplied to the boiler in a coal-fired power station, or to the combuster in a gas turbine, is rejected as waste heat. Parasitic losses; that is, energy consumed by the auxiliary plant such as pumps and fans, lower the overall power station efficiency still further. The best coal-fired stations rarely exceed overall efficiencies of 40 percent and a gas turbine operating in open-cycle mode will have an even lower value. The situation can be improved in the former case by recovering some of the heat rejected to the condenser as the steam from the turbine is condensed back to boiler-feed water. In the latter case, the exhaust gas from the gas turbine can be used to raise steam, which can be put to use in a process such as paper making or for community heating purposes. Efficiencies in this mode of operation can exceed 70 percent. The gas turbine can also be operated in so-called "combined-cycle" mode where the steam raised from the exhaust gas can be used to drive a steam turbo-alternator set to produce more electricity. Such an arrangement can have a fuel conversion efficiency of around 60 percent.

Intercontinental Transmission

As power supply systems in individual countries have become larger, high-capacity transmission circuits have been installed to interconnect national systems and allow the export and import of power. Typical of such systems is the cross-channel interconnector between England and France which has a capacity of 2000 megawatts and which allows the U.K. to take advantage of relatively cheap French nuclear power. Other recent examples of such interconnectors include those between the Channel Islands and France and between Scotland and Northern Ireland. Many of these are based on high-voltage direct current transmission via submarine cables, rather than AC transmission, which is typically employed on land-based overhead lines. Similar high-power interconnectors between the U.K. and Scandinavia were being planned as the century closed.

Environmental Regulation

In the latter decades of the century, there was increasing concern over the environmental impacts of electricity generation. These included acid rain from the sulfur dioxide emitted by coal combustion, and the safety and waste issues with nuclear power generation. Governments increasingly began to apply regulatory limits, requiring the installation of technologies such as gas scrubbing on power station exhaust streams. The desire to protect rivers in their natural state has also limited further growth in large hydroelectric power schemes in many countries. As the century closed, the link between carbon dioxide from fossil fuel combustion and global climate change was giving rise to regulation affecting all fossil fuel generation

and encouraging the development of renewable fuels for electricity generation.

See also **Electrical Power Distribution; Electricity Generation and the Environment; Energy and Power; Hydroelectric Power Generation; Turbines**

IAN BURDON

Further Reading

Edgerton, D. *Science, Technology and the British Industrial 'Decline' ca. 1870–1970*. Cambridge University Press/ Economic History Society, Cambridge, 1996.

Elliot, D. *Energy, Society, and Environment*, Routledge, London, 1997.

Hughes, T.P. *Networks of Power: Electrification in Western Society, 1880–1930*. Johns Hopkins University Press, Baltimore, 1893.

Leggett, J. *The Carbon War: Global Warming at the End of the Oil Era*. Penguin, London, 1999.

MacNeill, J.R., Ed. *Something New under the Sun: An Environmental History of the Twentieth Century*. Penguin, London, 2001.

Reddish, A. and Rand, M. The environmental effects of present energy policies, in *Energy, Resources and Environment*, Blunden, J. and Reddish, A., Eds. Hodder & Stoughton (in association with The Open University), London, 1991.

Wolf, D.E., Wolf, D.D. and Lowitt, R. *Big Dams and Other Dreams: The Six Companies Story*. University of Oklahoma Press, Norman, 1996.

Electrical Power Distribution

While the first commercial power station in San Francisco in 1879 was used for arc lighting (using a spark jumping a gap as the source of light) for street lamps, these had limited application. Edison's carbon filament lamp was the stimulus for the spread of electric lighting. A few of Edison's buildings and some private residences had their own generators, but Edison also recognized there was a need for a generating and distribution system. Edison's distribution system was first demonstrated in London, with a temporary installation running cables under the Holburn Viaduct in early 1882 that provided power for the surrounding district. The first permanent central electric generating station was Edison's Pearl Street Station in New York that went into operation in September 1882 and provided electricity (with a meter) to 85 customers in a 1 square mile (2.6 square kilometers) area. The Pearl Street Station used direct current (DC). In DC systems, the current flows in one direction, with a constant voltage. The dissipation of energy limits the size of DC systems and requires the source of electric generation to be close to the customer. Alternating current (AC) systems, in which the current changes direction (in today's public electricity supply, 50 or 60 times per second), overcame this limitation.

Transformers, invented by William Stanley Jr. in 1883, allowed voltage from an AC generator to be stepped up to high voltages for transmission and then back down again to lower voltage for utilization. At higher transmission voltages (and thus lower current), less electrical energy is lost. Power could thus be delivered efficiently over long distances and the source of generation could be far from the customers (in Edison's DC system, each generating system could only supply a few square kilometers or so). This would be particularly important once electrical energy was generated by hydropower stations, which were often many kilometers from the population or industry they served. In 1885 George Westinghouse, head of Westinghouse Electric Company, bought the patent rights to Nikola Tesla's AC polyphase system and placed its first power station in operation in 1886. The introduction of the AC induction motor in 1888 gave the AC system a great advantage in providing industry with electrical power. The AC system has been the dominant form of power transmission and distribution ever since.

Power transmission deals with the systems that move the power over very long distances from sources of generation such as hydro stations, fossil fuel plants, or nuclear power plants to central points of distribution called substations located in or near areas of large power demand such as cities and towns. In 1890, Westinghouse installed a 19-kilometer, 4000-volt transmission line from Willamette Falls to Portland. In 1897, the British Empire's first long-distance high-voltage transmission (11,000 volts or 11 kilovolts) of electric power traveled 27 kilometers from St-Narcisse on Batiscan River to Trois-Rivières, Québec, Canada. In 1903, the Shawinigan Water & Power Company installed the world's largest generator (5000 watts) and the world's largest and highest voltage line—136 kilometers and 50 kilovolts (to Montreal).

Today, the aspects of the electric power transmission system that are most apparent are the overhead transmission lines strung on the large steel towers visible in the countryside. These lines, called tower lines, transmit power over long distances. Each tower line has three power wires and two ground wires. The power wires carry the electric power and hang from insulators connected to the tower arms. The ground wires are connected at the top of the towers and protect the power carrying conductors from lightning strokes. Power

transmission occurs at high voltages; 500 or 750 kilovolts are not uncommon.

Multiple tower lines are connected together at a switching station. The connection occurs at a structure called a "bus." Between the tower line and the bus is a switch called a circuit breaker. The purpose of the circuit breaker is to disconnect the tower line from the bus in times of trouble, of which there are many types. Lightning hitting the tower line and causing a short circuit, power flow through a transmission line exceeding the power delivery capacity of the line, the wind blowing the transmission line into a tree—are all examples of transmission line troubles. If such an occurrence is not quickly isolated and fixed, it may damage the line or spread to other facilities. Devices called relays continually monitor the transmission line for signs of trouble. When the relay detects trouble on a transmission line, it signals circuit breakers to open. The opening of the circuit breaker then de-energizes the transmission line and clears the trouble. The power transmission system is made up of a complex arrangement of transmission lines and switching stations that provide many redundant paths from sites of generation to locations of customer demand.

This redundancy provides reliability. Since there are multiple redundant paths, if one transmission line should come out of service because of trouble, the power will flow over the remaining lines. Power system analysis is the discipline that insures that power transmission system will deliver power to the customer with sections of the transmission system out of service. This analysis places great demands on power system design engineers.

Prior to the advent of digital computers, power system analysis was performed on large analog computers called AC boards. The AC board was composed of miniature components that could be connected together to represent the power system. Engineers would spend many hours plugging in components to represent the power system to be studied and manually recording the power flows on data sheets. Digital computer programs were introduced in the early 1960s to solve the power flow problem. Early programs could only solve systems composed of a few hundred switching stations. Improvements in analytical techniques and computer speeds today permit the solution of 30,000 switching stations. The improvement in analytical techniques permitted the design of larger and more complicated interconnections. Today, most of the North America east of the Rocky Mountains is one large interconnected electric system. As the interconnection became larger and more complicated, it also became more difficult to operate. To insure reliable operation, computer systems called system control and data acquisition (SCADA) were introduced in the late 1960s to monitor and control the power transmission system.

A SCADA system is housed in a power company's control center, where it receives information about the operation of the transmission lines and circuit breakers. When the SCADA system detects an overloaded transmission line or a circuit breaker operation, it alerts the control room operator with an audible alarm. The alarms and system status are displayed on mimic boards, large electronic displays of the transmission power. Sophisticated computer programs provide the operator additional information including what operating moves will relieve overload facilities and what transmission flows will result if facilities trip out of service. These SCADA systems assist in providing high reliability of the transmission system in delivering power to the distribution system. Reliability of the transmission system is very important because transmission system breakdowns can cause widespread outages or blackouts.

The first widespread power outage in the U.S. was the Northeast Blackout that occurred in 1965—a relay failure blacked out 30 million people with over 20,000 megawatts of demand in the northeast U.S. and Canada. On January 3, 2001, India blacked out 220 million people in a large part of seven states for 13 hours. The most recent U.S. blackout occurred on August 14, 2003, when a huge failure blacked out New York City, Cleveland, Detroit, and Toronto affecting customers with a demand of 61,800 megawatts. On September 28, 2003, a huge blackout affected most of Italy. All of these blackouts were breakdowns in the power transmission system. Failures in the power distribution system are usually much more limited in scope.

The job of the power distribution system is to get the power from the substations to the customer. The aspects of electric power distribution that are most apparent are the wooden pole lines adjacent to streets and roads in towns and villages. The wires that carry the power from the distribution substation to the customer are connected to these poles with insulators. A pole line may contain primary circuits and secondary circuits. The primary circuit delivers the power at 4, 13, or 26 kilovolts from the substation to the pole top distribution transformer. A pole top distribution transformer resembles a large coffee can attached to the pole. The distribution transformer transforms the voltage down to the utilization or outlet

voltage, which in the U.S. is 120 volts and in the UK is 240 volts, well known to international travelers whose electric appliances may require adapters for use. The secondary circuit delivers the power from the distribution transformer to the customer. Unlike the redundant network grid of the transmission systems, the distribution system is usually radial—there is just one path for the power to flow from the distribution substation to the customer. If the path gets interrupted, the customer loses power. Since the path is comprised of overhead wires on a pole line, it is subject to damage by snow, ice, tree limbs, lightning, animals, and wind. Storms can cause significant damage to a distribution system. To improve the reliability of the distribution system, the overhead lines are often equipped with sectionalizing and tie switches. These switches permit damaged parts of the system to be isolated and the power supplied by alternate paths. In more heavily populated areas, the distribution system may be placed underground. Sometimes the underground insulated wires are placed in concrete ducts and sometime they are directly buried. The duct system is more reliable, but it is also more expensive.

Re-Emergence of DC Power Distribution

With the development in the 1970s of high voltage thyristor valves that functioned as AC/DC converters it became possible to transmit DC power at high voltages (high voltage direct current or HVDC) over large distances. HVDC is particularly suitable for linking interconnected systems and for submarine links. The cross-Channel link installed between Britain and France in 1986 is one example. In the modern era of deregulated electricity supplies, cross-border energy trading between nations whose grids are synchronously interconnected can exploit the time difference of peak load periods on national systems.

See also **Electric Motors; Electrical Energy Generation and Supply, Large Scale; Electricity Generation and the Environment; Energy and Power**

PETER VAN OLINDA

Further Reading

Bowers, B., Ed. *An Early History of Electricity Supply: The Story of the Electric Light in Victorian Leeds*. Peregrinus, Stevenage, 1986.

Davis, L.J. *Fleet Fire: Thomas Edison and the Pioneers of the Electric Revolution*. Arcade, New York, 2003.

Electrical Transmission and Distribution: Reference Book. Westinghouse Electric Corporation, East Pittsburgh, 1964.

Grainger, J. *Power System Analysis*. McGraw-Hill, New York, 1994.

Harrison, K. *Currents of the Brazos: An Illustrated History of Brazos Electric Power Cooperative, Inc 1941–1991*. Nortex Press, Austin, 1991.

Jonnes, J. *Empires of Light: Edison, Tesla, Westinghouse, and the Race to Electrify the World*. Random House, 2003.

Miller, R. *Power System Operation*. McGraw-Hill Professional, New York, 1994.

Platt, H. *The Electric City: Energy and the Growth of the Chicago Area 1880–1930*. University of Chicago Press, Chicago, 1991.

Warkentin, D. *Electric Power Industry: In Nontechnical Language*. Pennwell, Tulsa, OK, 1998.

Useful Websites

Electric Power from Niagara Falls: http://ublib.buffalo.edu/libraries/projects/cases/niagara.htm

History of the U.S. Power Industry, 1882–1991: http://www.eia.doe.gov/cneaf/electricity/page/electric_kid/append_a.html

U.S. Department of Energy: http://tonto.eia.doe.gov/FTPROOT/electricity/056296.pdf

A Brief History of Public Power in the Pacific Northwest: http://www.ppcpdx.org/Side_Bar/history.htm

Mimic Board: http://www.ee.washington.edu/energy/apt/nsf/previous/MimicBoard.htm

Power system: http://www.howstuffworks.com/blackout.htm

High Voltage Direct Current Transmission Systems Technology Review Paper : http://www.worldbank.org/html/fpd/em/transmission/technology_abb.pdf

Electricity Generation and the Environment

Fossil fuel thermal generating technologies were a mainstay of both twentieth century electricity generation and environmental attention. While concern with declining urban air quality, initially at the center of this attention, dated back to the nineteenth century, it was the substantial post-World War II rise in electricity consumption that resulted in the later prominence of these concerns. The impacts of fossil fuel extraction and transportation were also a source of significant twentieth century environmental attention, but concern over atmospheric emissions dominated. Although initial concern focused on particulate emissions, attention shifted to acidic emissions from the 1970s onward, and the final decade of the century was dominated by concern with the impact of fossil fuel emissions on climate. This later concern with fossil fuel greenhouse gas emissions, primarily thermally produced carbon dioxide (CO_2) but also fugitive emissions (i.e., not caught by a capture system) such as methane from coal seams and from gas extraction and distribution systems, reinforced an increasing emphasis on alternative generating

technologies. Some of these, notably macro-hydro and nuclear fission, were significant twentieth century technologies in their own right, and their environmental impacts are briefly discussed below. However, as the twentieth century closed this emphasis was increasingly turning to renewable energy technologies and the potential for significant further efficiencies in both electricity generation and consumption, including the drive to "decarbonize" electricity generation by turning away from fossil fuel technologies.

Coal-fired thermal generation was pivotal to the rising twentieth century concern with poor urban air quality. Coal contains many impurities, and the resulting particulate, and typically acidic, emissions were a major source of urban air pollution. Heavily populated and industrialized countries such as Britain were particularly affected, and London's "Great Smog" of December 5–9, 1952, is regarded as a turning point in pollution control. This event, leading to an estimated 4000 deaths, provided impetus for the U.K.'s Clean Air Acts of 1956 and 1968, which focused on curbing particulate emissions. Power station emissions were controlled at the source by both particulate removal and the use of taller emission stacks to promote dispersal over wider areas. The electrostatic precipitators used to remove particulates soon became a standard power station design feature, while the use of taller emission stacks inadvertently resulted in the issue of invisible pollutants as a significant regional concern. This concern centered on the environmental acidification resulting from the sulfur dioxide (SO_2) and to a lesser degree nitrous oxide (NO_x) emissions from coal as well as oil-fired thermal generation. Particularly marked in Europe, where British emissions were deposited on Norway and Sweden, and in North America, where U.S. emissions were similarly deposited in Canada, these concerns resulted in a milestone in pollution control, the UN Economic Commission for Europe's 1979 Geneva Convention on Long-Range Transboundary Air Pollution. While these concerns remained the focus of legislative attention throughout the rest of the twentieth century and resulted in the widespread installation of flue gas desulfurization technology, it was CO_2 emissions that soon came to dominate attention.

As the twentieth century progressed and the scale of hydroelectric installations increased they became the focus of considerable concern. Massive projects such as the "New Deal" Tennessee Valley project in the U.S. and the postwar Australian Snowy Mountains project were on a scale previously unseen, and many of their environmental impacts (such as changed hydrological regimes) only came to light many decades later. Such macro-hydro projects, particularly in the developing world, came to be seen as symbols of environmental devastation.

While the first civil nuclear power plants of the 1950s were greeted with an optimism encapsulated by the phrase "too cheap to meter," this optimism subsided as the impact and implications of the entire nuclear fuel cycle became clearer. Public disquiet was dominated by concerns with operating safety, the implications of which were graphically demonstrated by the Ukrainian Chernobyl accident of 1987, but it was the broader considerations of nuclear waste storage, decommissioning, and their economic costs that placed a political question mark over the future of nuclear technology.

Rising environmental concerns from the 1960s and the oil price shocks of the 1970s resulted in greater interest in the development of renewable energy technologies. Concern with climate change, institutionalized in the UN Framework Convention on Climate Change (1992) and the subsequent Kyoto Protocol mandating greenhouse gas emission reductions, significantly reinforced this concern in the 1990s. Photovoltaic cells, used for spacecraft from the 1950s, became a focus of terrestrial attention from the 1970s. By the end of the century both cell efficiency increases and cost decreases were dramatic, but generating costs still exceeded those of fossil fuels by a factor of around 4. The great success story of this period was wind power. Although the first 1 megawatt-plus demonstration wind turbines the 1970s overextended contemporary materials technology, turbines of more than 400 kilowatts were in widespread commercial use by the 1990s. By the close of the century, wind power was not only cost competitive with fossil fuels but was also the fastest growing of all electricity generating technologies. In addition, with increasingly sophisticated materials technologies, 1 megawatt-plus turbines were entering the commercial market. Electricity generation from waste and biomass was also increasingly common, while other renewable options such as wave and ocean power, micro-hydro, and geothermal technologies were gaining increasing attention.

The push to decarbonize electricity generation in the 1990s also resulted in proposals for fuel switching, particularly to cleaner and less carbon-intensive natural gas, which offered vastly improved generating efficiencies. Other fossil options such as cleaner coal technologies and sequestering of greenhouse gases were also pur-

sued. Combined heat and power, in which the waste heat from thermal generation is productively captured, also became more attractive in applications that required heat inputs such as industry and domestic heating. The potential for significant energy savings from both the supply and demand sides of technological and behavioral change was widely discussed in commercial/industrial and political circles, although by the close of the century little progress had been made on this front.

See also **Biomass Power Generation; Hydroelectric Power Generation; Power Generation, Recycling; Solar Power Generation; Technology, Society, and the Environment; Wind Power Generation**

STEPHEN HEALY

Further Reading

Boehmer-Christiansen, S. and Skea, J. *Acid Politics: Environmental and Energy Policies in Britain and Germany*. Belhaven Press, London, 1991.

Elliot, D. *Energy, Society and Environment*. Routledge, London, 1997.

Goldemberg, J. *Energy, Environment and Development*. Earthscan, London, 1996.

McNeill, J. Air pollution since 1900, in *Something New Under the Sun: An Environmental History of the Twentieth Century World*, McNeill, J., Ed. Penguin, London, 2000.

McNeill, J. Energy regimes and the environment, in *Something New Under the Sun: An Environmental History of the Twentieth Century World*, McNeill, J., Ed. Penguin, London, 2000.

Reddish, A. and Rand, M. The environmental effects of present energy policies, in *Energy, Resources and Environment*, Blunden, J. and Reddish, A., Eds. Hodder & Stoughton (in association with The Open University), London, 1991.

Electrocardiogram (ECG)

The electrocardiogram (ECG or EKG) is a graphic measurement of the electrical activity of the heart produced by an electrocardiograph, or ECG machine. ECGs are used to assess heart function and are especially important in the detection of electromechanical abnormalities such as arrhythmias and myocardial infarctions (blockages in the small vessels of the heart, commonly termed a heart attack). The development of the electrocardiograph has been closely linked with advances in the science of electrophysiology, the study of the role of electricity in the functioning of living organisms. As early as 1842, the Italian physicist Carlo Matteucci demonstrated that each heart beat is accompanied by an electrical discharge. The following year, the German physicist Emil Dubois-Raymond confirmed Matteucci's findings, coining the term "action potential" to denote the electrical charge accompanying muscular contraction. In 1856, Rudolph von Koelliker and Heinrich Muller noted a two-fold electrical charge accompanying heart systole, which is the muscular contraction of the heart to pump blood through the blood vessels of the body. Identification of this electrical charge presaged the later division of the heart's electrical signature into a series of so-called waves and complexes. Practical application of these findings to the medical field of cardiology (study of the heart) depended on the creation of a reliable means of (1) measuring the electrical activity of the heart and (2) relating the measurements to the functional anatomy of the heart (i.e., establishing the electrical signatures denoting structure, size, blood supply, electrical pathways, and rhythm of the heart). A first significant step in this direction was the creation of the capillary electrometer by the French physicist Gabriel Lippmann in 1872. This was a thin glass tube of mercury and sulfuric acid that registered variations in electrical potential that could be observed under a microscope. The capillary electrometer was used in 1878 by the British physiologists John Burden Sanderson and Frederick Page to confirm the two phases of the heart's electrical activity, later designated the QRS complex and T wave. The first human electrocardiogram produced using the capillary electrometer was published by the British physiologist Augustus D. Waller in 1887.

However, the tracing or graph produced by capillary electrometry was still crude. This was due in part to the sluggishness of the measuring needle that required a mathematical correction and in part because a crucial component of modern electrocardiography—the use of multiple leads to produce a more sophisticated picture of the heart's electrical activity—had yet to be invented. The increase in the number of measuring leads used in electrocardiography along with the corresponding degree of accuracy came in the three stages. In 1891, British physiologists William Bayliss and Edward Starling improved the capillary electrometer by adding a second lead to the right hand to the original lead over the heart itself. Using this technique, they discovered a third phase of the heart's electrical activity, which was later designated the P wave. A second stage in this development was the Dutchman Willem Einthoven's creation of a triangular arrangement of leads. His arrangement, termed Einthoven's triangle, established what have come to be known as the standard

leads in electrocardiography, designated I, II, and III. First used in 1912, this system has continued to be the basis of the ECG lead arrangement.

Einthoven, a major figure in the development of modern electrocardiography, distinguished five phases of the heart's electrical action, noting that the middle pulse was in fact a complex with three separate components. Einthoven dubbed the five separate deflections of the electrocardiogram the P, Q, R, S, and T waves, terminology that is still used. Einthoven's second major contribution to the development of electrocardiography was his use of a new technology, the string galvonometer. Einthoven first employed the string galvonometer in 1901, and the following year he published the first human ECG using it. Einthoven played a key role in the production of the string galvonometer for commercial use, and the first was sold in 1908 to Edward Schafer at the University of Edinburgh. Just as important as his technical and commercial innovation was Einthoven's theoretical contribution to understanding the electrophysiology of the heart. In 1906, he published the first systematic presentation of normal and abnormal ECG tracings, thus describing a number of important electrophysiological abnormalities of the heart. Einthoven was awarded the Nobel Prize in medicine in 1924 for the invention of the electrocardiograph.

The modern arrangement of 12 ECG leads was pioneered in 1942 by Emanuel Goldberger. Although recent experiments with computerized ECGs employ 15 or more leads, the 12-lead arrangement remains standard. Other important developments involved improvements in power source and portability. The successful use of vacuum tubes was reported in 1928, marking a shift from the mechanical to the electric electrocardiograph. The first practical direct-writing ECG "cart" was developed in the 1950s. Advances in computerization, beginning in 1959 with the creation of the first analog-to-digital converter designed specifically for the electrocardiograph, resulted in the standardization by the mid-1980s of ECG carts capable of generating a computerized interpretation of ECG results. Despite the rise of new technologies such as echocardiography, which uses ultrasound to examine the heart, the ECG remains a vital tool of clinical cardiology and is likely to remain so for some time because of its simplicity, cost-effectiveness, and accuracy.

See also **Cardiovascular Disease, Diagnostic Methods; Diagnostic Screening**

TIMOTHY S. BROWN

Further Reading

Burch, G.E. and DePasquale, N.P. *A History of Electrocardiography*. Year Book Medical Publishers, Chicago, 1964.

Bynum, W.F., Lawrence, C. and Nutton, V., Eds. *The Emergence of Modern Cardiology*. Wellcome Institute for the History of Medicine, 1985.

Coumel, P. *Electrocardiography, Past and Future*. New York Academy of Sciences, New York, 1990.

Fye, W.B. A history of the origin, evolution, and impact of electrocardiography. *Am. J. Cardiol.*, 73, 937–949, 1994.

Horan, L.G. The quest for optimal electrocardiography. *Am. J. Cardiol.*, 41, 126–129, 1978.

Rautaharju, P.M. A hundred years of progress in electrocardiography 1: early contributions from Waller to Wilson. *Can. J. Cardiol.*, 3, 362–374, 1987.

Snellen, H.A. *Willem Einthoven (1860–1927) Father of Electrocardiography*. Kluwer, Dordrecht, 1995.

Useful Websites

Jenkins, D. 12-lead ECG library homepage. Electrocardiogram (ECG, EKG) Library, Llandough Hospital, Cardiff, Wales, 1999: http://homepages.enterprise.net/djenkins/ecghome.html.

Electrochemistry

Electrochemistry deals with the relationship between chemical change and electricity. Under normal conditions, a chemical reaction is accompanied by the liberation or absorption of heat and not of any other form of energy. However, there are many so-called electrochemical reactions that when allowed to proceed in contact with two electronic conductors joined by conducting wires, will generate electrical energy in this external circuit. Current between the electrodes (usually metallic plates or rods) is carried by electrons, while in the electrolyte, a nonmetallic ionic compound either in the molten condition or in solution in water or other solvents, ions carry the current.

Conversely, the energy of an electric current can be used to bring about many chemical reactions that do not occur spontaneously. The process in which electrical energy is directly converted into chemical energy is called electrolysis. The products of an electrolytic process have a tendency to react spontaneously with one another, reproducing the substances that were reactants and were therefore consumed during the electrolysis. If this reverse reaction is allowed to occur under proper conditions, a large proportion of the electrical energy used in the electrolysis can be regenerated. This possibility is used in accumulators or storage cells, sets of which are known as storage batteries.

Electrochemistry owes its rapid developments in the nineteenth century to the invention of the battery by Alessandro Volta in 1800, which provided the first source of continuous current. Within six weeks of Volta's report, two English scientists, William Nicholson and Anthony Carlisle, used a chemical battery to discover electrolysis, producing hydrogen and oxygen by passing an electric current through water. This is how the science of electrochemistry began. By 1809 the English chemist Humphry Davy had used a more powerful battery to isolate several very active metals for the first time—sodium, potassium, calcium, strontium, barium, and magnesium—by electrolyzing their liquid compounds. Michael Faraday, who was Davy's assistant at the time, studied electrolysis quantitatively and showed that the amount of energy needed to separate a gram of a substance from its compound is closely related to the atomic weight of the substance. However, the relationship between electrical charge and energy was not understood until the electron was discovered in 1896. The generation of an electrical current from gaseous hydrogen and oxygen in contact in an acidic electrolyte was first demonstrated by William Grove in 1839.

During the 1820s, Davy and Faraday attached pieces of zinc or iron to the copper sheeting on the bottom of ships, in order to prevent saltwater corrosion. Galvanic corrosion of metals occurs when two dissimilar metals are in contact in a conducting solution (such as seawater). An electrochemical cell forms, whereby the more active metal becomes the anode and corrodes (by losing metal ions) and the cathode is protected. This technique was not seriously used until 1956 when the U.S. Navy began to experiment with platinum-clad titanium anodes for the protection of their ships and submarines. Today pipelines and oilrigs are also electrochemically protected by placing zinc or magnesium "sacrificial anodes" near the steel structure.

In the twentieth century many new applications of electrochemical reactions were developed. Technologies of generation of electrical power from batteries and fuel cells are described in separate entries. Here we describe the use of electrolysis in metallurgy, electroplating and surface finishing, and analytical chemistry.

Most technologically important metals, except iron and steel, are either obtained or refined by electrolytic processes. In 1886, Charles Martin Hall in the U.S. and Paul-Louis-Toussaint Heroult in France independently and simultaneously discovered the modern method of commercially producing aluminum by electrolysis of purified alumina (Al_2O_3) dissolved in molten cryolite (Na_3AlF_6) ore. Copper is refined by electrolysis in aqueous copper sulfate solutions. Unrefined copper, with its impurities is made the anode, and thin sheets of pure copper form the cathode. As the anodes corrode away, pure metal is deposited on the cathodes. The widespread use of electrolysis for depositing protective coats on electronic conductors dates back to the middle of the twentieth century. In electroplating, the object to be coated is made the cathode in an electrolytic cell. Titanium, alkaline earth, and alkali metals are obtained by electrodeposition from molten salts, and automobile parts are chrome plated to protect them from corrosion. Surface finishing is the reverse of electroplating. Instead of coating or plating, metal is removed from the surface, leaving a smooth, clean finish.

In the chemical industry, electrolysis of seawater to obtain caustic soda (sodium hydroxide) and chlorine as a byproduct, used since 1900, has become one of the largest volume productions. Since the 1980s electrochemical process have been developed for the synthesis of a variety of inorganic compounds to the production of such synthetic fibers as nylon. Novel structures such as multilayers, nanowires, and granular composites can be formed or modified.

Electroanalysis employs electrochemical phenomena for quantitative analysis, for example detection and determination of various aqueous and gaseous ions. In electrochemical sensors the gas to be detected diffuses through a semipermeable membrane to the working electrode. At the electrode the gas is either oxidized or reduced, and the resulting current provides the signal. Electroanalytical techniques have been used for environmental monitoring and industrial quality control since the late 1960s, and more recently for biomedical and pharmaceutical analysis.

Certain technical advances in the 1980s and 1990s—such as the development of ultramicroelectrodes, the design of tailored interfaces and molecular monolayers, the coupling of biological components and electrochemical transducers, the synthesis of ionophores and receptors containing cavities of molecular size, and the development of high-resolution scanning probe microscopes—led to a substantial increase in the utility of electrochemical reactions. By the close of the century, electrochemical reactions were used in a vast range of processes and applications in multiple scientific fields and industries.

See also **Batteries, Primary and Secondary; Fuel cells**

MUZAFFAR IQBAL

Further Reading

Bard, A.J. *Electrochemical Methods: Fundamentals and Applications*. Wiley, New York, 2001.

Bockris, J.O. *Modern Electrochemistry*. Plenum, New York, 2001.

Brett, C.M.A. *Electroanalysis*. Oxford University Press, Oxford, 1998.

Bruce, Peter G., Ed. *Solid State Electrochemistry*. Cambridge University Press, Cambridge, 1997.

Dubpernell, G. and Westbrook, J.H., Eds. *Selected Topics in the History Of Electrochemistry*. Electrochemical Society, Princeton, 1978.

Ostwald, W. *Electrochemistry: History and Theory*. Verlag von veit, Leipzig, 1895. Translated by Date, N. P., for the Smithsonian Institution and the National Science Foundation, Amerind, New Delhi, 1980.

Samuel, R. *The Founders of Electrochemistry*. Dorrance, Philadelphia, 1975.

Stock, J.T., Ed. *Electrochemistry, Past and Present*. American Chemical Society, Washington D.C., 1989.

Taillefert, M. and Rozan, T.F., Eds. *Environmental Electrochemistry: Analysis of Trace Element Biogeochemistry*. American Chemical Society, Washington D.C. Distributed by Oxford University Press, 2002.

Wang, J. *Analytical Electrochemistry*. Wiley, New York, 2000.

Useful Websites

Electrochemistry Encyclopedia: http://electrochem.cwru.edu/ed/encycl/

Electroencephalogram (EEG)

An electroencephalogram (EEG) is the graphic measurement and analysis of electrical activity in the brain. The technique allows diagnostic information about brain activity to be determined noninvasively; that is, without recourse to surgery. Electrical activity in the body was studied from as early as the 1780s as seen in Luigi Galvani's work on "animal electricity." A professor at the university in Bologna, Galvani demonstrated that contact with an electric circuit would cause frogs' legs to contract. Richard Caton, an English physician, was the first scientist to report that he had recorded electrical activity in the brains of rabbits and monkeys. Caton used a galvanometer to detect the electrical signals. Galvanometers work on the principle that a current-carrying conductor in a magnetic field experiences a force. The conductor is mounted on pivots in a constant magnetic field. Once current flows through the needle, the needle deflects, with the deflection proportional to the electric current. In his experiments, Caton detected small currents of varying direction when two points were placed on the external surface of the skull.

The first human EEG was recorded by a German physician, Dr. Hans Berger (1873–1941), who published his findings in 1929. Berger used two sheets of aluminum foil as electrodes, placing one at the back and one at the front of the head. The signals were detected with a Siemen's double coil galvanometer. Electrical activity in the brain can be measured from the scalp surface due to volume conductor effects. The signal obtained will depend on the distance from the electrodes to the source of the signal and on the intervening material. The electrodes are separated from the brain by the skull and meninges (membranes) lining the brain, and large areas of the brain are comparatively remote from the electrodes. It is estimated that recordings of electrical activity through scalp electrodes can sample activity from only 0.1 to 1 percent of the neuronal population.

Although the basic measurement principles in electroencephalography are similar to those of electrocardiography, the amplitudes of the EEG potentials are approximately 1000 times smaller. Electroencephalography technology was slower to develop because of the difficulty in measuring and amplifying small signals. In the 1930s valve amplifiers were used to boost the signals detected because amplifiers increased the amplitude of a signal while maintaining the signal integrity. Pen writers were introduced in the 1940s as the recording device to graph the EEG potentials. The pen writers were driven by the signal output, and the pen moved in response to the signal changes. The pens were placed over a moving paper chart in order to record amplitude variations along a time axis.

Electroencephalography enabled clinicians to obtain diagnostic information about brain activity and brain abnormalities because EEG signals represent collective neural activity in the brain. In the 1930s EEGs found immediate application in exploring epileptic seizure disorders. Epilepsy is characterized by pronounced EEG abnormalities. American scientists Alfred L. Loomis, E.N. Harvey, and G.A. Horbart were among the first to study human sleep EEGs and EEG patterns for different stages of sleep. Walter (1936) moved EEG research into the study of other disorders with the discovery of patterns associated with brain tumors. Brain tumors can be located by a careful examina-

tion of EEG potentials along the whole contour of the scalp. A number of common EEG waveforms were subsequently characterized. Alpha waves are characteristic of the awake closed eyes state. They have amplitudes of 20–50 microvolts and a frequency of 8.0–13.5 Hz. They are not recorded if the eyes are opened. Theta waves have a frequency of 7.5–4 Hz and occur in drowsiness. Delta waves have a frequency of less than 4Hz and occur in subjects who are in deep sleep or coma. Mu waves are associated with motor activity, and lambda waves with viewing patterned visual displays

The basic measurement method has seen little change since the 1950s. Electronic amplifiers replaced valve amplifiers, which increased sensitivity and provided better signal to noise ratios. Electrodes were made smaller and placed at many points over the scalp. Electrode positioning, or montage, was standardized to ensure that each electrode was placed at the same position on the head with each subject. The 10/20 montage system defines positions on the scalp in relation to particular points on the head such as the "nosion" at the front of the head and the "inion" at the back of the head. Electrodes are placed at specified fractions of distance between these positions. The electrodes detect local differences in electrical activity between electrodes placed on the scalp. The responses are amplified by a type of electronic amplifier designed to amplify the difference between two signals. The responses are displayed as numerous graphs, each representing potential differences between electrodes in the 10/20 montage. Systems generally have 8 or 16 channels, each recording EEG data. The standard presentation is a graph of voltage as it varies with time. The EEG is displayed on monitors or printed by chart recorders.

As the frequency is also important in characterizing signals, researchers began looking at the frequency spectrum. In 1965 J. W. Cooley and J. W. Tookey introduced the use of the "fast Fourier transform" to EEG as the basis of power spectral analysis. Frequency analysis also introduced new ways of displaying frequency information as color-coded patterns. The colored patterns depict the frequency distribution of electrical activity in different areas of the brain.

The 1970s saw the development of an associated technology called evoked potential measurement. Evoked potential studies examined sensory function by recording brain responses to a specific sensory stimulus. For example visual responses may be stimulated in the subject who is viewing changing patterns on a monitor. Detection of the *evoked* response among all the other brain activity that occurs simultaneously is achieved through a signal-processing technique called signal averaging. The stimulus is repeated, and, over a short poststimulus period, responses are monitored. In the poststimulus period, activity not related to the stimulus will be randomly distributed. The visual evoked responses, which are time-locked to the stimulus, will occur at the same point in relation to the stimulus. If sufficient responses are recorded, the evoked response can be distinguished from other electrical brain activity once the responses are averaged.

By the end of the twentieth century, interest in EEG measurements had waned. Computed tomography (CT) and magnetic resonance imaging (MRI) provided high-quality brain images, and functional MRI applications were developed which offered new diagnostic and research possibilities. In the 1980s, EEG brain topography was developed. This technique exploits the availability of a great number of digitized channels providing information simultaneously from an array of points across the scalp. Developments with evoked potentials led to research into their use for establishing depth of anesthesia for patients undergoing surgery. EEGs have continued to be used, however, in conjunction with positron emission tomography (PET) and MRI in studies that merge information from numerous modalities to provide more comprehensive information about brain activity.

See also **Neurology; Nuclear Magnetic Resonance (NMR, MRI); Positron Emission Tomography (PET); Tomography in Medicine**

COLIN WALSH

Further Reading

Brown, B.H., Smallwood, R.H., Barber, D.C., Lawford, P.V. and Hose, D.R. *Medical Physics and Biomedical Engineering*. Institute of Physics Publishing, U.K., 1999.

Cooper, R., Ossetton, J.W., Shaw, JC. EEG Technology, 3rd edn. Butterworth–Heinemann, 1980.

Davidovits, P. *Physics in Biology and Medicine*. 2nd edn. Harcourt/Academic Press, U.S.

Dyro, F.P *The EEG Handbook*. Little, Brown & Company, U.S., 1989.

Neidermeyer, E. *Historical Aspects (of EEG)*, in *EEG: Basic Principles, Clinical Applications and Related Fields*, 3rd edn., Neidermeyer, E. and Da Silva, F.L., Eds. Williams & Wilkins, Baltimore, 1987.

Useful Websites

http://www.neuro.com/megeeg/
http://www.epub.org.br/cm/n03/tecnologia/historia.htm

Electronic Communications

The development of digital computing and communication technology in the 1940s and 1950s was largely driven by Cold War military needs in the midst of closed-world politics. Extensive funding was provided for large-scale research and development projects during this period by the U.S. military. The origins of the communication of digital information can be attributed to the Whirlwind computer, which was developed under these conditions at the Massachusetts Institute of Technology (MIT) in Boston. This was a powerful general-purpose digital computer orientated toward real-time control and flight simulation. However, it eventually found use as the control computer in the semiautomatic ground environment (SAGE) air defense system. SAGE connected remote early-warning radar stations in the far reaches of the Arctic with control centers in the heartland to automatically direct fighter aircraft to intercept the perceived onslaught of Soviet bombers carrying doomsday nuclear arsenals. A means had to be invented to communicate digital information over long distances between the radar stations and the SAGE control centers, and this resulted in the techniques for the long-distance communication of digital data being developed. This is the origin of the modem (as it later became known) and the system became operational in 1952. The name is derived from the process of the modulation and demodulation, whereby a waveform of digital data (ones and zeros) are superimposed onto a sinusoidal carrier wave, since the square waveform of digital data cannot be sent over distances. This information is then extracted with the process of demodulation on the receiver side. It should also be noted that many other key computing technological developments resulted from the SAGE project, such as video displays, magnetic core memory, and networking among others. This technology diffused into civilian use with the SABRE airlines reservation system built by IBM for American Airlines in 1964, which inherited a lot of technological developments from SAGE. The system used modems to transmit data signals over ordinary analog telephone channels and was an early example of the general trend of the diffusion of military-sponsored computing technology into broader society.

Among other commercial data communications systems was the remote teletype terminal, built in the late 1950s. Modems are measured by the rate of data transmission in bit per second (bps or baud) and these first commercial implementations provided 110 bps. Bell Laboratories later developed the first commercial component modems, which were largely used for remote terminals and for the long-distance interconnection of computers for data transfer. By 1962 the first commercial modems were on the market. The Bell 103 was the first modem to provide full duplex transmission and had a data transmission speed of 300 bps. Further theoretical signal processing developments in optimization of encoding and compression techniques increased the speed of data transmission, another theme which would continue until the end of the century. The Bell 212 modem later provided 1200 bps. However, an important development was the setting of internationally agreed standards, with the International Telephone and Telegraph Consultative Committee (CCITT) establishing the V.21 standard in 1964, which defined 300 bps communication protocols. Handshaking—which begins before data transmission, establishes an electrical path and synchronization, and enables the two devices to send messages agreeing on a communications protocol. Communications protocols define message format and rate of transmission. Having the data communication industry produce modems to mutually agreed standards was a key factor in later proliferation, although the Bell standards continued to be used and supported.

Computing technology grew at a phenomenal pace during the 1960s with the mainframe and later the minicomputer proliferating into universities, research centers, and large corporations globally. The need for these machines and user communities to communicate and share resources led to concept of a "net," which was first discussed and proposed in 1961 by Leonard Kleinrock at MIT. His work covered much of the essential theory of data traffic behavior. Packet switching data networks as a way of achieving this was a significant shift in thinking from traditional data communications. This network would interconnect geographically dispersed computing resources, utilizing modem communications over long-distance telephone lines. The stream of digital data would be broken into packets, which were sent one at a time with no central control. Although there were research networks in the U.K. and elsewhere, the U.S. Advanced Research Projects Agency packet switching network (ARPANET) was conceived by Paul Baran at the Rand Corporation in early 1960s, and was the first to be implemented in 1969 in the U.S. Computers were connected to the network via switching nodes which were called interface message processors (IMPs) which were

themselves computers and connected to each other via permanent (leased) telephone lines and adaptive equalizer modems supporting up to 56,000 bps (or 56 kbps). Communication tasks were conceptually dealt with in logical layers of functionality and utility, abstracting higher levels of application communication for lower levels of physical communication of data. This meant that exactly how the data was transported became irrelevant to an application.

The Advanced Research Projects Agency Network (ARPANET) was designed primarily for resource sharing. However, it came to be used in ways that were unintended, which arguably drove its success. Electronic messaging (e-mail, net notes, or mail) was the most significant activity on the early ARPANET. Although messaging had been present on host computers from as early as 1965, in 1971 programmer Ray Tomlinson from the company Bolt, Beranek and Newman (BBN), which designed and built the IMPs, adapted earlier code to automatically relay messages between different host computers interconnected on the net. This text message file was transmitted using the underlying file transport mechanisms. He also introduced the "username@hostname" format of messaging addressing which was later to become the standard addressing scheme for e-mail. His system was later refined and incorporated into the standard ARPANET interconnection package, and e-mail quickly became the network's most used feature, generating more traffic over the network than any other application. E-mail also opened up ways of communication that enhanced the collaboration of remotely located research groups, and was used socially between nontechnical users in ways totally unanticipated and unforeseen by the creators of the ARPANET.

However, access to ARPANET was restricted to an elite at the top research universities and institutions in the U.S. and elsewhere to a limited extent. In response to this, a separate network emerged in the later 1970s: Usenet. This was an electronic messaging system built in 1979 at Duke University, North Carolina, with no formal funding. Graduate students programmed their Unix computer to automatically link up with other Unix systems. They used homemade modems and standard Unix programs (UUCP) to build a system that automatically dialed up over ordinary telephone lines and replicated files between a series of computers. Usenet had no formal structure, and was available to all who were interested as long as they had access to the Unix operating system, which was available at low cost to the academic and computer research community. Connections later became available via personal computers.

Usenet did not only distribute e-mail; it also facilitated a structured group discussion system called "netnews," later known as "newsgroups." These emulated an interactive newspaper, but allowed the reader to be selective. More importantly, readers could participate in discussion threads, interactively contributing to the flow of thoughts and information. The Unix developers at AT&T's Bell Labs provided a lot of support for Usenet, since they saved millions of dollars on development by using the Usenet community for help with debugging their software. This in turn enabled the network to further proliferate, since Bell Labs provided a lot of resources that helped with the distributional logistics of the network. Digital Equipment Corporation (DEC) also supported Usenet, and ultimately the spread of Usenet and Unix encouraged the sale of computers that ran the Unix operating system. The Usenet newsgroups communities provided considerable technical help for peers in an ever-proliferating range of topics that extended well beyond that of technical concerns. Usenet grew rapidly, with the number of sites and articles per day doubling yearly until 1988. Once again the scale of this growth took the original Usenet community by surprise since it was not at all anticipated. Although there was hesitation about this growth from the original developers, seemingly insurmountable problems were investigated and solved by the Usenet community. This community referred to Usenet as 'the Net'. Many communities in the U.S. and around the world established "Free Nets," based on free Usenet software. Later in the 1980s the popularity of the mass-produced personal computers and modems meant that Usenet became more accessible to individuals and small organizations via dial-up UUCP connections. Some of these Free Nets had access to the worldwide newsgroups of Usenet. A gateway through to the ARPANET was established in 1981 at The University of California at Berkeley.

Bitnet was another early network that provided a form of electronic messaging. Most notably it was the origins of the list server, a facility that provided one-to-many messaging. Once again it was entirely independent of the ARPANET developments and was primarily linked to IBM computers at academic installations on the East Coast of the U.S. It was initiated in 1981. The list server was also a very important factor in the discussion of new ideas for the improvement of the network as well as resolution and acceptance of new stan-

dards. It could thus be considered self-perpetuating, a characteristic generally thought to be unique to the (later) Internet. The Listserv program was moved over to the Unix operating system in 1991, freeing it from the proprietary hold of the IBM mainframe. The Listserv communities ultimately migrated to the Internet once it emerged, where proliferation took place on a much grander scale, for the same reasons as they did on the original Bitnet network.

The development of the personal computer (PC) in the mid-1970s by hobbyists was to have profound impact on electronic communications in broader society. PCs were made possible by integrated circuit technology, which was a spin-off of microelectronics used in 1960s aerospace projects such as the Apollo Lunar Program. By the 1980s PCs and modems had become consumer items. The Hayes modem, invented in 1977 by Dennis Hayes, provided autonomous functionality since the modem itself contained a computer in the form of a microprocessor. Thus it was able to perform "intelligent" functions such autodial and autoanswer automatically and could be programmed to respond in certain ways to different line conditions and communication protocols. Many low-level communication tasks were "off-loaded" to the modem. A consequence of this was that it allowed personal computers to be set up as central communication hosts, running bulletin board software (BBS), which provided discussion forums and electronic messaging among other more utilitarian functionality such as file transfer. Electronic messaging communities developed outside of the corporations and academic institutions and reflected diverse social interest groups, increasing the diffusion of the concept of electronic messaging and communication by attracting like-minded people to join these communities and participate. Gateways to other BBSs as well as other established networks such as Bitnet further fueled growth and acceptance of PC-based messaging. This included many independent (and in most cases noncommercial) networks such as FidoNet, DASNet and others.

Large commercial messaging systems soon followed which also became economically feasible with the popularity of the PC in the early to mid 1980s. Most of these were located in the U.S. and made extensive use of long-distance commercial data services to act as a "backbone" between distributed hosts (large computers) that were coupled together to appear logically as one system. Users connected directly to local access points via the standard modems and telephone lines, and generally were charged for the amount of time they were connected. The systems were all proprietary based on the functionality of the BBS as described above, and included CompuServe, The Source, BIX, WELL, Dialcom and many others. Some only offered e-mail, like MCI Mail and Telemail. Gateways between these systems and the Internet (as the ARPANET became in the mid1980s) and other public networks were introduced in time. A limitation of these systems was the technical barriers in establishing a connection with them. Each system required connection with unique modem communication parameters such as number of bits, parity, and checksum with each system having its own unique configuration. This required a technical literacy from users, and acted as a barrier to broader proliferation; however, many users were successful and brought together millions of users who became familiar with communicating via electronic messaging and accessing information "online" from their homes and workplaces.

It is worth noting similar developments of electronic communities in Europe and elsewhere. The Prestel system originated from work done in the 1960s by the U.K. Post Office on a standard they called Viewdata. It was an interactive data service that initially required dedicated terminals with all the electronics built and was unique in that it provided limited graphics. It was initially designed to coexist with a television set as the video display unit. The service was based on host computers on which subscribers could publish information. There were gateways to other information service providers. It was later possible to access the Prestel service from a personal computer and modem with 75 bps upstream and 300 bps downstream communication. Although it was popular, it was not commercially successful. This is in contrast to the French Minitel system (initially known as Télétel), which was similar in concept except that the French telephone network gave the Minitel Terminals away to subscribers to reduce the cost of printing telephone directories. The service was highly successful. In 1987 it was the world's largest e-mail system and 6.5 million terminals were in use by 1995.

These disparate systems and electronic communities discussed above ultimately converged and migrated onto the Internet in the 1990s, once that emerged as the dominant network. Most of the technology was software and could be "reprogrammed," drawing these communities of users together onto one system that nobody but everybody owned. For this and other reasons, the global proliferation of the Internet acted as the great

unifying force for the growth of global electronic messaging or "cyber" communities, with e-mail, newsgroups and forums as the basis of communications. Many of the communities discussed above rapidly migrated to similar means of communication of the Internet, propelling its phenomenal growth.

Not all digital communication technology was successful. The telecommunication carrier and service provider community attempted to convert the subscriber analog telephone system in a deterministic way to a digital circuit with the Integrated Services Digital Network (ISDN) system. The I.120 ISDN recommendation was published in 1984 by the Comité Consultatif International Téléphonique et Télégraphique (CCITT), but it was unspecific in areas and open to interpretation. Despite extensive promulgation and marketing, ISDN did not find widespread use or success. Reasons included incompatibility between manufacturers' implementations, as well as higher bandwidth coming from analog modem developments by the time the ISDN services were working. Thus there was little incentive for users to convert to ISDN, although it did find some success later as a data link backup for permanent digital network lines.

The series of CCITT modem standards cumulated in the 56 kbps V.90 standard, finalized in 1998, although successful manufacturer-specific implementations were being used years before that. The introduction of digital subscriber lines (DSL) late in the 1990s provided a significant increase bandwidth (512 kbps). However having such "broadband" bandwidth available did not change the way users used the net, rather it merely speeded up the process, and opened up further avenues of use that users embraced, such as the transfer of digital images, sound and video files. It should also be noted that the advantages of high bandwidth were not always supported further up the network than the point of access. DSL also had line distance restrictions, and thus was restricted to urban areas.

The broad use of digital electronic message communications in most societies by the end of the century can be attributed to a myriad of reasons. Diffusion was incremental and evolutionary. Digital communication technology was seeded by large-scale funding for military projects that broke technological ground, however social needs and use drove systems in unexpected ways and made it popular because these needs were embraced. Key technological developments happened long before diffusion into society, and it was only after popularity of the personal computer that global and widespread use became commonplace. The Internet was an important medium in this regard, however the popular uses of it were well established long before its success. Collaborative developments with open, mutually agreed standards were key factors in broader diffusion of the low-level transmission of digital data, and provided resistance to technological lock-in by any commercial player. By the twenty-first century, the concept of interpersonal electronic messaging was accepted as normal and taken for granted by millions around the world, where infrastructural and political freedoms permitted. As a result, traditional lines of information control and mass broadcasting were challenged, although it remains to be seen what, if any, long-term impact this will have on society.

See also **Computer Networks; Computers, Personal; Internet; Packet Switching; World Wide Web**

BRUCE GILLESPIE

Further Reading

Abbate, J. *Inventing the Internet*. MIT Press, Cambridge, MA, 1999.

Bijker, W., Hughes, T., Pinch, T., Eds. *The Social Construction of Technological Systems*. MIT Press, Cambridge, MA, 1989.

Ceruzzi, P.E. *A History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Edwards, P.N. *The Closed World*. MIT Press, Cambridge, MA, 1996.

Grier, D. and Campbell, M. A social history of Bitnet and Listserv, 1985–1991. *Ann. Hist. Comp.*, 22, 2, 32–41, 2000.

Hauben, M. *Netizens, On the History and Impact of the Usenet and the Internet*. Institute of Electrical and Electronics Engineers Computer Society, Los Alamitos, 1997.

Hughes, T.P. *Rescuing Prometheus*. Vintage Books, New York, 2000.

Quarterman, J.S. *The Matrix, Computer Networks and Conferencing Systems Worldwide*. Digital Press, Boston, 1990.

Winston, B. *Media Technology and Society, A History: From the Telegraph to the Internet*. Routledge, London, 1998.

Electronics

Electronic systems in use today perform a remarkably broad range of functions, but they share the technical characteristic of employing electron devices such as vacuum tubes, transistors, or integrated circuits. Most electron devices in use today function as electric switches or valves, controlling a flow of electrons in order to perform

useful tasks. Electron devices differ from ordinary electromechanical switches or current- or voltage-control devices in that an applied electric current or field controls electron flow rather than a mechanical device. Electronic devices are "active," like machines, but have no moving parts, so engineers distinguish them both from electromechanical devices and from other "passive" electrical components such as wires, capacitors, transformers, and resistors. When the word electronics was coined around 1930, it usually referred to so-called vacuum tubes (valves), which utilize electrons flowing through a vacuum. With the advent of the transistor in the late 1940s, a second term emerged to describe this new category of "solid-state" electron devices, which performed some of the same functions as vacuum tubes but consisted of solid blocks of metal.

Few of the basic tasks that electronic technologies perform, such as communication, computation, amplification, or automatic control, are unique to electronics. Most were anticipated by the designers of mechanical or electromechanical technologies in earlier years. What distinguishes electronic communication, computation, and control is often linked to the instantaneous action of the devices, the delicacy of their actions compared to mechanical systems, their high reliability, or their tiny size.

The electronics systems introduced between the late nineteenth century and the end of the twentieth century can be roughly divided into the applications related to communications (including telegraphy, telephony, broadcasting, and remote detection) and the more recently developed fields involving digital information and computation. In recent years these two fields have tended to converge, but it is still useful to consider them separately for a discussion of their history.

The origins of electronics as distinguished from other electrical technologies can be traced to 1880 and the work of Thomas Edison. While investigating the phenomenon of the blackening of the inside surface of electric light bulbs, Edison built an experimental bulb that included a third, unused wire in addition to the two wires supporting the filament. When the lamp was operating, Edison detected a flow of electricity from the filament to the third wire, through the evacuated space in the bulb. He was unable to explain the phenomenon, and although he thought it would be useful in telegraphy, he failed to commercialize it. It went unexplained for about 20 years, until the advent of wireless telegraphic transmission by radio waves. John Ambrose Fleming, an experimenter in radio, not only explained the Edison effect but used it to detect radio waves. Fleming's "valve" as he called it, acted like a one-way valve for electric waves, and could be used in a circuit to convert radio waves to electric pulses so that that incoming Morse code signals could be heard through a sounder or earphone.

As in the case of the Fleming valve, many early electronic devices were used first in the field of communications, mainly to enhance existing forms of technology. Initially, for example, telephony (1870s) and radio (1890s) were accomplished using ordinary electrical and electromechanical circuits, but eventually both were transformed through the use of electronic devices. Many inventors in the late nineteenth century sought a functional telephone "relay"; that is, something to refresh a degraded telephone signal to allow long distance telephony. Several people simultaneously recognized the possibility of developing a relay based on the Fleming valve. The American inventor Lee de Forest was one of the first to announce an electronic amplifier using a modified Fleming valve, which he called the Audion. While he initially saw it as a detector and amplifier of radio waves, its successful commercialization occurred first in the telephone industry. The sound quality and long-distance capability of telephony was enhanced and extended after the introduction of the first electronic amplifier circuits in 1907. In the U.S., where vast geographic distances separated the population, the American Telephone and Telegraph Company (AT&T) introduced improved vacuum tube amplifiers in 1913, which were later used to establish the first coast-to-coast telephone service in 1915 (an overland distance of nearly 5000 kilometers).

These vacuum tubes soon saw many other uses, such as a public-address systems constructed as early as 1920, and radio transmitters and receivers. The convergence of telephony and radio in the form of voice broadcasting was technically possible before the advent of electronics, but its application was greatly enhanced through the use of electronics both in the radio transmitter and in the receiver.

World War I saw the applications of electronics diversify somewhat to include military applications. Mostly, these were modifications of existing telegraph, telephone, and radio systems, but applications such as ground-to-air radio telephony were novel. The pressing need for large numbers of electronic components, especially vacuum tubes suitable for military use, stimulated changes in their design and manufacture and contributed to improving quality and falling prices. After the war, the expanded capacity of the vacuum tube industry

contributed to a boom in low-cost consumer radio receivers. Yet because of the withdrawal of the military stimulus and the onset of the Great Depression, the pace of change slowed in the 1930s. One notable exception was in the field of television. Radio broadcasting became such a phenomenal commercial success that engineers and businessmen were envisioning how "pictures with sound" would replace ordinary broadcasting, even in the early 1930s. Germany, Great Britain, and the U.S. all had rudimentary television systems in place by 1939, although World War II would bring nearly a complete halt to these early TV broadcasts.

World War II saw another period of rapid change, this one much more dramatic than that of World War I. Not only were radio communications systems again greatly improved, but for the first time the field of electronics engineering came to encompass much more than communication. While it was the atomic bomb that is most commonly cited as the major technological outcome of World War II, radar should probably be called the weapon that won the war. To describe radar as a weapon is somewhat inaccurate, but there is no doubt that it had profound effects upon the way that naval, aerial, and ground combat was conducted. Using radio waves as a sort of searchlight, radar could act as an artificial eye capable of seeing through clouds or fog, over the horizon, or in the dark. Furthermore, it substituted for existing methods of calculating the distance and speed of targets. Radar's success hinged on the development of new electronic components, particularly new kinds of vacuum tubes such as the klystron and magnetron, which were oriented toward the generation of microwaves. Subsidized by military agencies on both sides of the Atlantic (as well as Japan) during World War II, radar sets were eventually installed in aircraft and ships, used in ground stations, and even built into artillery shells. The remarkable engineering effort that was launched to make radar systems smaller, more energy efficient, and more reliable would mark the beginning of an international research program in electronics miniaturization that continues today. Radar technology also had many unexpected applications elsewhere, such as the use of microwave beams as a substitute for long-distance telephone cables. Microwave communication is also used extensively today for satellite-to-earth communication.

The second major outcome of electronics research during World War II was the effort to build an electronic computer. Mechanical adders and calculators were widely used in science, business, and government by the early twentieth century, and had reached an advanced state of design. Yet the problems peculiar to wartime, especially the rapid calculation of mountains of ballistics data, drove engineers to look for ways to speed up the machines. At the same time, some sought a calculator that could be reprogrammed as computational needs changed. While computers played a role in the war, it was not until the postwar period that they came into their own. In addition, computer research during World War II contributed little to the development of vacuum tubes, although in later years computer research would drive certain areas of semiconductor electron device research.

While the forces of the free market are not to be discounted, the role of the military in electronics development during World War II was of paramount importance. More-or-less continuous military support for research in electronic devices and systems persisted during the second half of the twentieth century too, and many more new technologies emerged from this effort. The sustained effort to develop more compact, rugged devices such as those demanded by military systems would converge with computer development during the 1950s, especially after the invention of the transistor in late 1947.

The transistor was not a product of the war, and in fact its development started in the 1930s and was delayed by the war effort. A transistor is simply a very small substitute for a vacuum tube, but beyond that it is an almost entirely new sort of device. At the time of its invention, its energy efficiency, reliability, and diminutive size suggested new possibilities for electronic systems. The most famous of these possibilities was related to computers and systems derived from or related to computers, such as robotics or industrial automation. The impetus for the transistor was a desire within the telephone industry to create an energy-efficient, reliable substitute for the vacuum tube. Once introduced, the military pressed hard to accelerate its development, as the need emerged for improved electronic navigational devices for aircraft and missiles.

There were many unanticipated results of the substitution of transistors for vacuum tubes. Because they were so energy efficient, transistors made it much more practical to design battery powered systems. The small transistor radio (known in some countries simply as "the transistor"), introduced in the 1950s, is credited with helping to popularize rock and roll music. It is also

worth noting that many developing countries could not easily provide broadcasting services until the diffusion of battery operated transistor receivers because of the lack of central station electric power. The use of the transistor also allowed designers to enhance existing automotive radios and tape players, contributing eventually to a greatly expanded culture of in-car listening. There were other important outcomes as well; transistor manufacture provided access to the global electronics market for Asian radio manufacturers, who improved manufacturing methods to undercut their U.S. competitors during the 1950s and 1960s. Further, the transistor's high reliability nearly eliminated the profession of television and radio repair, which had supported tens of thousands of technicians in the U.S. alone before about 1980.

However, for all its remarkable features, the transistor also had its limitations; while it was an essential part of nearly every cutting-edge technology of the postwar period, it was easily outperformed by the older technology of vacuum tubes in some areas. The high-power microwave transmitting devices in communications satellites and spacecraft, for example, nearly all relied on special vacuum tubes through the end of the twentieth century, because of the physical limitations of semiconductor devices. For the most part, however, the transistor made the vacuum tube obsolete by about 1960.

The attention paid to the transistor in the 1950s and 1960s made the phrase "solid-state" familiar to the general public, and the new device spawned many new companies. However, its overall impact pales in comparison to its successor—the integrated circuit. Integrated circuits emerged in the late 1950s, were immediately adopted by the military for small computer and communications systems, and were then used in civilian computers and related applications from the 1960s. Integrated circuits consist of multiple transistors fabricated simultaneously from layers of semiconductor and other materials. The transistors, interconnecting "wires," and many of the necessary circuit elements such as capacitors and resistors are fabricated on the "chip." Such a circuit eliminates much of the laborious process of assembling an electronic system such as a computer by hand, and results in a much smaller product. The ability to miniaturize components through integrated circuit fabrication techniques would lead to circuits so vanishingly small that it became difficult to connect them to the systems of which they were a part. The plastic housings or "packages" containing today's microprocessor chips measure just a few centimeters on a side, and yet the actual circuits inside are much smaller. Some of the most complex chips made today contain many millions of transistors, plus millions more solid-state resistors and other passive components.

While used extensively in military and aerospace applications, the integrated circuit became famous as a component in computer systems. The logic and memory circuits of digital computers, which have been the focus of much research, consist mainly of switching devices. Computers were first constructed in the 1930s with electromechanical relays as switching devices, then with vacuum tubes, transistors, and finally integrated circuits. Most early computers used off-the-shelf tubes and transistors, but with the advent of the integrated circuit, designers began to call for components designed especially for computers. It was clear to engineers at the time that all the circuits necessary to build a computer could be placed on one chip (or a small set of chips), and in fact, the desire to create a "computer on a chip" led to the microprocessor, introduced around 1970. The commercial impetus underlying later generations of computer chip design was not simply miniaturization (although there are important exceptions) or energy efficiency, but also the speed of operation, reliability, and lower cost. However the inherent energy efficiency and small size of the resulting systems did enable the construction of smaller computers, and the incorporation of programmable controllers (special purpose computers) into a wide variety of other technologies. The recent merging of the computer (or computer-like systems) with so many other technologies makes it difficult to summarize the current status of digital electronic systems. As the twentieth century drew to a close, computer chips were widely in use in communications and entertainment devices, in industrial robots, in automobiles, in household appliances, in telephone calling cards, in traffic signals, and in a myriad other places. The rapid evolution of the computer during the last 50 years of the twentieth century was reflected by the near-meaninglessness of its name, which no longer adequately described its functions.

From an engineering perspective, not only did electronics begin to inhabit, in an almost symbiotic fashion, other technological systems after about 1950, but these electronics systems were increasingly dominated by the use of semiconductor technology. After virtually supplanting the vacuum tube in the 1950s, the semiconductor-based transistor became the technology of choice for most

subsequent electronics development projects. Yet semiconducting alloys and compounds proved remarkably versatile in applications at first unrelated to transistors and chips. The laser, for example, was originally operated in a large vacuum chamber and depended on ionized gas for its operation. By the 1960s, laser research was focused on the remarkable ability of certain semiconducting materials to accomplish the same task as the ion chamber version. Today semiconductor devices are used not only as the basis of amplifiers and switches, but also for sensing light, heat, and pressure, for emitting light (as in lasers or video displays), for generating electricity (as in solar cells), and even for mechanical motion (as in micromechanical systems or MEMS).

However, semiconductor devices in "discrete" forms such as transistors, would probably not have had the remarkable impact of the integrated circuit. By the 1970s, when the manufacturing techniques for integrated circuits allowed high volume production, low cost, tiny size, relatively small energy needs, and enormous complexity; electronics entered a new phase of its history, having a chief characteristic of allowing electronic systems to be retrofitted into existing technologies. Low-cost microprocessors, for example, which were available from the late 1970s onward, were used to sense data from their environment, measure it, and use it to control various technological systems from coffee machines to video tape recorders. Even the human body is increasingly invaded by electronics; at the end of the twentieth century, several researchers announced the first microchips for implantation directly in the body. They were to be used to store information for retrieval by external sensors or to help deliver subcutaneous drugs. The integrated circuit has thus become part of innumerable technological and biological systems.

It is this remarkable flexibility of application that enabled designers of electronic systems to make electronics the defining technology of the late twentieth century, eclipsing both the mechanical technologies associated with the industrial revolution and the electrical and information technologies of the so-called second industrial revolution. While many in the post-World War II era once referred to an "atomic age," it was in fact an era in which daily life was increasingly dominated by electronics.

See also **Audio Recording; Computers, Uses and Consequences; Control Technology—Electronic Signals; Electronic Communications; Integrated Circuits; Lasers; Lighting Techniques; Radio Receivers; Radio Transmitters; Rectifiers; Transistors; Valves/Vacuum Tubes**

David Morton

Further Reading

Dummer, G.W.A. *Electronic Inventions and Discoveries*, 4th edn. Institute of Physics Publishing, Bristol, 1997.

Fagen, M.D. et al., Eds. A History of Engineering and Science in the Bell System, 5 vols. 1975–84.

Riordan M. and Hoddeson, L. *Crystal Fire: The Birth of the Information Age*. W.W. Norton, New York, 1998.

Electrophoresis

Electrophoresis is a separation technique that involves the migration of charged colloidal particles in a liquid under the influence of an applied electric field. The word is derived from *electro*, referring to the energy of electricity, and *phoresis*, from the Greek verb *phoros*, meaning "to carry across." Electrophoresis has many applications in analytical chemistry, particularly biochemistry. It is one of the staple tools in molecular biology and it is of critical value in many aspects of genetic manipulation, including DNA studies, and in forensic chemistry.

Swedish biochemist Arne Tiselius carried out studies on proteins and colloids in the 1920s, and in 1930 introduced electrophoresis as a new technique for separating proteins in solution on the basis of their electrical charge. Tiselius was awarded the 1948 Nobel Prize in chemistry for this work, and the technique became a common tool in the 1940s and 1950s. Biological molecules such as amino acids, peptides, proteins, nucleotides, and nucleic acids, possess ionizable groups. At any given pH (concentration of hydrogen ions), these molecules exist in solution as electrically charged species either as cations (positive, or +) or anions (negative, or −). Depending on the nature of the net charge, the charged particles will migrate either to the cathode or to the anode. For example, proteins in an electric field separate according to size, shape, and charge with charges contributed by the side chains of the amino acids composing the proteins. The charge of the protein depends on the hydrogen ion content of the surrounding buffer with a high ionic strength resulting in a greater charge.

The information derived from electrophoretic methods has long been considered particularly valuable because of the gentleness of the method. The application of other separation methods,

involving precipitation or aggressive chemicals, may easily cause damage to substances as unstable as biocolloids. In electrophoretic separation, the migrating substances remain in the same medium during the entire process, and the observation method is likely to give warning of eventual irreversible changes accompanying the separation.

If the material under investigation is in the form of a reasonably stable, dilute suspension or emulsion containing microscopically visible particles or droplets, then electrophoretic behavior can be observed directly. Information relevant to soluble material can also be obtained if the substance is adsorbed on to the surface of a carrier. The term zone electrophoresis refers to electrophoresis that is carried out in a supporting medium, whereas moving boundary electrophoresis is carried out entirely in a liquid phase. Most electrophoretic methods use a supporting media, such as starch powder, paper, polyacrylamide gel, or agar gel. Paper was used in the late 1940s and in 1955, Oliver Smithies used starch gel as a medium to minimize convection; unfortunately however, starch has its own small charged groups, which produce an electroendosmotic flow. In 1959 Leonard Ornstein and Baruch Joel Davis, and independently, Samuel Raymond and L. Weintraub introduced the use of polyacrylamide gel. Stallan Hjertén showed that polyacrylamide gels could act as "molecular sieves," whose pore size allowed separation of proteins by size even if their charge was the same. Polyacrylamide gels are still among the most used for electrophoretic separation of proteins, agarose gels are used for nucleic acids.

Moving boundary electrophoresis, the type first perfected in 1937 by Tiselius in separating the similar components of blood serum, has largely been superseded by simpler and less expensive methods. The original method involves dialyzing a buffered (i.e., constant pH) solution of the material under investigation in a large U-shaped tube with electrodes at each end. Separation of the proteins could be observed by observing light that was deflected by refractive index gradients at a boundary. A weakness of the method is that only partial separations can be achieved; full resolution of individual components is not possible due to diffusion and heat-driven convection.

Zone electrophoresis, developed by Stellan Hjertén in Sweden in 1967, involves particles that are supported on a relatively inert and homogenous solid or gel framework in order to minimize diffusion and convectional disturbances, and thus improve separation. Zone offers many advantages over moving boundary electrophoresis, achieving complete separation of all electrophoretically different components while the introduction of a stabilizing medium permits the use of much simpler and less expensive equipment. Perhaps most importantly, much smaller samples can be studied than in moving boundary electrophoresis, which has significant implications for DNA analysis as well as criminal detection. When filter paper is used as a medium for low voltage electrophoresis, paper strips are dipped in buffer solution and clamped between electrodes. When the separation is complete (up to 20 hours), the paper strips are removed from the electrophoresis tank and dried in an oven. The separated components are then typically located by staining with a dye that binds to the proteins.

Zone electrophoresis is the preferred form of analysis among scientists. Developments in zone electrophoresis have closely paralleled chromatography with both techniques sharing a number of supporting media and methods for estimating the separated components. However, it remains easier to separate proteins with the zone method.

Gel electrophoresis, the zone form employed with DNA, uses the frictional resistance of a gel to separate nucleic acids and proteins. Higher current densities can be used than with paper. A hot gel mixture is poured into a casting tray to assume a desired shape as it polymerizes. DNA is loaded on to the gel and electricity is applied for about 20 minutes. After staining, the separated macromolecules in each lane can be seen in a series of bands spread from one end of the gel to the other. The successive application of two separation steps in perpendicular directions (two dimensions) is known as 2-D gel electrophoresis. The first step separates proteins by charge. The second step separates the proteins by via their molecular weight. Since the mid-1980s, gel electrophoresis has been used to create genetic fingerprints for forensic or courtroom use by identifying particular DNA molecules. The number and position of bands formed on each lane of gel is the actual genetic "fingerprint" of that DNA sample. Viral DNA, plasmid DNA, and particular segments of chromosomal DNA can all be identified in this way. Another use is the isolation and purification of individual fragments containing interesting genes, which can be recovered from the gel with full biological activity. Using this technology, it is possible to separate and identify protein molecules that differ by as little as a single amino acid.

In the 1970s Rauno Virtanen was the first to use buffer-filled stationary narrow bore tubes (capil-

laries) to stabilize against convection. In the early 1980s James Jorgenson and Krynn Lukacs used much narrower capillaries, and developed analytic packages to simplify handling of data. Detection is usually by ultraviolet (UV) absorbance or coupling to mass spectrometers.

Electrophoresis on Earth is limited to very lightweight materials. For separation of minerals, for example, Earth's gravity causes convection currents, as well as gravitational settling. The space shuttle has been used to perform experiments in zero gravity in a free fluid, for example to produce small amounts of an electrophoretically purified protein. No commercial applications are yet developed.

See also **Chromatography; X-Ray Crystallography**

CARYN E. NEUMANN

Further Reading

Audubert, R. and de Mende, S. *The Principles of Electrophoresis*. Macmillan, New York, 1960.
Bruno, T.J. *Chromatographic and Electrophoretic Methods*. Prentice Hall, Englewood Cliffs, NJ, 1991.
Issaq, H.J. *Century of Separation Science*. Marcel Dekker, New York, 2001.
Laitinen, H.A. and Ewing, G.W., Eds. *A History of Analytical Chemistry*. American Chemical Society, New York, 1977, pp. 324–325.
Shaw, D.J. *Electrophoresis*. Academic Press, London, 1969.
Wieme, R. J. *Agar Gel Electrophoresis*. Elsevier, Amsterdam, 1965.

Encryption and Code Breaking

The word cryptography comes from the Greek words for "hidden" (*kryptos*) and "to write" (*graphein*)—literally, the science of "hidden writing." In the twentieth century, cryptography became fundamental to information technology (IT) security generally. Before the invention of the digital computer at mid-century, national governments across the world relied on mechanical and electromechanical cryptanalytic devices to protect their own national secrets and communications, as well as to expose enemy secrets. Code breaking played an important role in both World Wars I and II, and the successful exploits of Polish and British cryptographers and signals intelligence experts in breaking the code of the German Enigma ciphering machine (which had a range of possible transformations between a message and its code of approximately 150 trillion (or 150 million million million) are well documented.

In many respects the construction of the Enigma was like other rotor cipher devices in use in Europe and America in the 1930s, such as the Hagelin machine invented in the 1920s. At the heart of Enigma was a set of three or more electromechanical rotors and a movable alphabet ring inscribed with letters indicating the rotor positions. However, the Enigma distinguished itself with its reflecting rotor, which caused plaintext (the original message), once scrambled, to pass through the machine a second time in the opposite direction, resulting in double encipherment. British mathematician Alan Turing and other cryptanalysts working at the super-secret Bletchley Park broke the encipherment with several tools, including mistakes made by German users of the machine (procedural errors such as sending the same message twice or using standard message formats too often), and with the fact that a particular letter could never encipher as itself. By 1942, Allied code breakers deciphered almost 4000 German communiqués each day. By some estimates the Bletchley Park effort to break Enigma shortened World War II by two years.

Figure 1. Marine version of an Enigma ciphering machine invented by Germany during World War II.
[*Courtesy of the photographer, Morton Swimmer, and the Bundesamt für Sicherheit in der Informationstechnik*.]

American military planners in the twentieth century—and especially in the Cold War—regarded superior information as a powerful "force multiplier" making up for the perennial problem of numerically inferior manpower. For most of the century, however, America's information advantage could be measured mainly in terms of secrecy rather than espionage. This was because spy agencies like the Army Cipher Bureau (established in 1917), the Navy Cryptanalytic Group (in 1924), and the National Security Agency (in 1952) virtually ignored in their massive number-crunching activities the issue of integrity—the corruption of data by mistakes, inflation, or enemy deception.

The key document pointing out this problem was American Bell Labs researcher Claude Shannon's "*Communication Theory of Secrecy Systems*," published in 1949. Shannon put the study of cryptography on firmer scientific ground by linking it to a formal information theory for reliable communications under noisy, error-prone conditions. His key insight was in showing that secrecy systems are "almost identical with a noisy communication system." Noise in the form of radio static, television snow, or background chatter, he argued, is very much like the process of encipherment. Cryptanalysis, conversely, involved the isolation of the signal (plaintext) from the noise (cipher key).

Public citizens in the Western world became much more engaged by the personal privacy implications of cryptographic national security in the 1950s and 1960s. Counterintelligence, they discovered, could be used not only against criminals, fascist dictators, and communists, but also against the counterculture youth movement and other political dissenters. Counterintelligence projects initiated against civil rights organizations and Vietnam War protesters for instance—which included authorized wiretapping of phones and electronic eavesdropping in homes and offices—outraged large numbers of freethinking individuals who felt that privacy is something everyone is entitled to as a right. Until about 1990, individuals had no access to the high levels of cryptography that were enjoyed by governments.

Both the problem and solution to cryptographic control were seemingly apparent in the multiply redundant and commercial interconnection of computers over telephone networks in the 1960s and 1970s. The need to address the issue of encryption to ensure privacy and security in electronic communications was highlighted by the unique enticements of networked computing as well as the extraordinary potential for malfeasance. Data security in information systems rapidly grew into a formal area of study within hardware and software engineering. Security experts realized that indirect losses—image problems as well as civil and criminal liabilities—were also a possible consequence of breaches in security. New network protective measures and access controls were added. Costs and benefits of particular security measures were calculated and recalculated.

However, public networks and computer time-sharing also allowed "every man at the console." The extraordinary movement of computational power into public hands contributed to a decentralized view of authority that competed with the surviving impetus for cryptographic national security. As Kenneth Dam and Herbert Lin have argued, "The broadening use of computers and computer networks [and cell phones] has generalized the demand for technologies to secure communications down to the level of individual citizens and assure the privacy and security of their electronic records and transmissions." Since the 1970s a growing global community of programmers ascribing to the vague hacker ethic that free information-sharing is a powerful positive good have competed with other interests motivated instead by profits of large corporations, security in electronic commerce, or the individual's demand for privacy.

Since the late 1970s, the popularization and commercialization of public-key, or "asymmetric," encryption technologies such as Pretty Good Privacy (PGP, available as freeware or low-cost commercial versions) and RSA (from the initials of Ron Rivest, Adi Shamir and Len Adleman, professors at Massachusetts Institute of Technology who developed a similar public-key approach) have weakened the hold of national governments on ownership of cutting-edge cryptographic technology. Martin Hellman, a professor at Stanford University, and two graduate students Whitfield Diffie and Ralph Merkle, discovered public-key encryption in 1976. (In 1997 it was disclosed that three employees of the British government had discovered the same approach several years earlier but had kept it a secret for reasons of national security.) Unlike classic private-key, or "symmetrical," encryption—where the same key is used to both encrypt and decrypt a message—Diffie and Hellman proposed a scheme to split the key. A *public key* would permit others to send the owner encrypted messages and verify the authenticity of messages received; a *private key*

would be held only by the owner and used to decrypt messages encrypted with the public key or to sign new messages with their personal "digital signature." Because the public key could not unlock messages encrypted with the public key, that key could be distributed widely without danger of interception by a third party, a danger implicit in symmetrical encryption.

A private key (whether for encryption or signature) consists of just two very large prime numbers, and the matching public key is those two numbers multiplied together. The security arises from the fact that whereas multiplying two large numbers together is quite easy (if tedious), given just the public key there is no practical way to find out its two factors, and without doing that one cannot break the code. For example, the old Data Encryption Standard, or DES, adopted by the National Institute of Standards and Technology (NIST) in 1977 used keys that were 56 digits long, which was secure at the time. By the 1990s it was recognized that those numbers were not sufficiently large enough to be impervious to those with special hardware or distributed computing (a few days number crunching). Encryption with a security key of 128 bits length (commonly used in electronic commerce in the 2000s) increased the "strength" of the encryption by a factor of 2^{72}.

Almost immediately after Diffie and Hellman's proposal, computer, networking, and telephone companies began developing public-key technology for secure voice encryption in digital telephony. The specter of widespread use of public-key cryptography with keyspaces (the collection of all possible keys for a given cryptosystem) so large that they could never be completely checked by brute force and its possible use by organized crime or foreign governments convinced the American government to propose "key escrow." Key escrow would allow legal wiretaps and the easy decryption of intercepted digital messages with a "back door" key held in trust by the NIST and the Department of the Treasury. The government dubbed its key-escrow technology the Clipper chip (announced in 1993) and advised manufacturers of soon-to-be-released telephone security devices to replace their own security schemes with Clipper. Though Clipper was never implemented for various technical and political reasons, the key escrow idea ignited a debate by civil libertarians, so-called "cryptoactivists," government authorities and lawmakers, and public interests that has yet to subside at the dawn of the twenty-first century.

PHILIP L. FRANA

Further Reading

Blanchette, J.-F. and Johnson, D.G. Cryptography, data retention, and the panopticon society. *Comp. Soc.*, 28, 1–2, 1998.

Dam, K.W. and Lin, H.S., Eds. *Cryptography's Role in Securing the Information Society*. National Academy Press, Washington D.C. 1996.

Denning, D.E. To tap or not to tap, *Commn. Asso. Comput. Mach.*, 36, 26–33, 1993.

Diffie, W. and Landau, S. Privacy on the Line: The Politics of Wiretapping and Encryption. MIT Press, Cambridge, MA, 1998.

Hinsley, F.H. and Stripp, A. *Codebreakers: The Inside Story of Bletchley Park*. Oxford University Press, Oxford, 1994.

Kahn, D. The Codebreakers: The Story of Secret Writing. Macmillan, New York, 1967.

King, H.R. Big Brother, the holding company: a review of key-escrow encryption technology. *Rutgers Comp. Tech. Law J.*, 21, 224–62, 1995.

Levy, S. *Crypto: How the Code Rebels Beat the Government: Saving Privacy in the Digital Age*. Viking, New York, 2001.

Rosenheim, S.J. *The Cryptographic Imagination: Secret Writing from Edgar Poe to the Internet*. Johns Hopkins University Press, Baltimore, 1997.

Shannon, C.E. Communication theory of secrecy systems. *Bell Syst. Tech. J.*, 28, 656–715, 1949.

Simmons, G.J., Ed. *Secure Communications and Asymmetric Cryptosystems*. Westview Press, Boulder, CO, 1982.

Useful Websites

Electronic Privacy Information Center: http://www.epic.org

Energy and Power

At the close of the twentieth century, electricity was so commonplace that it would be difficult to imagine an existence without light, heat, and music bowing to our command at the flick of a switch. Children who could barely stretch high enough to toggle a light switch now have dominion over phenomena that less than a century ago would have been considered inconceivable. The temptations of such power have proved hard to resist.

The repercussions of the rapacious appetite for control of energy among Western industrial nations have not been confined to the lot of the individual, however. As in previous eras, when the control of mechanical or biological power carried financial, geographical, and social significance, the use and abuse of electrical energy now additionally carries environmental, political, and moral implications. Developments in energy and power in the twentieth century must therefore be considered within these broader thematic areas as the genera-

tion and consumption of energy are inextricably linked with practically the whole spectrum of human existence.

At the beginning of the twentieth century, despite the fact that many components of modern electronics such as the battery had already been invented 100 years earlier, body power was still the norm, especially in rural areas. Horses, carriages, tow paths, water mills, and the like were the standard means of transport and power for a large proportion of the population, despite the growth of electricity and the 130 supply companies that were operating by 1896 in Britain. Even in urban settings, only lighting and telegraphy were advanced to the stage where the benefits were generally enjoyed as a result of Thomas Edison's invention of the light bulb in 1879 and Alexander Graham Bell's first telephone transmission in 1878.

By 1900 in Britain the main features of an electricity supply industry had been established. The system was based on the generation of high-voltage alternating current (AC), with transformers stepping down voltages for local use. However, one obstacle that the industry had to overcome was the lack of standardization across local areas. In some parts, direct current (DC) equipment was still installed, and local voltage levels and frequencies varied considerably. Despite problems posed by these variations, at the start of the century most of the appliances that are now taken for granted had appeared. Space heaters, cookers, and lighting equipment were not yet in every home, but the very speed at which their use was adopted was testament to the flexibility and popularity of electricity. In 1918 electric washing machines became available, and in 1919 the first refrigerator appeared in Britain. They had already been introduced for domestic use in the U.S. in 1913. Electricity had been firmly accepted as the energy of the future. Demand from the residential sector started to boom and spurred further research. Most importantly, perhaps, by the 1920s in Britain the domestic immersion heater began to take over the duties of coal. The use of electric trolleys and trains, which had been running since the end of the nineteenth century, also continued to expand, and underground travel developed swiftly. Electricity also made advances in communications possible, from the telegraph and the telephone, to the broadcasting boom of the 1920s. In 1928 the construction of a British national grid system began, and it took less than ten years before the system was in operation. This alacrity is partly to be explained by the influence of World War I. The war's heavy demands on manufacturing acted as a great incentive for the rapidly evolving electricity industry, particularly with regard to improving the efficiency of supply. Thereafter, the rebuilding and expansion of industry across the industrialized world began. In Russia, Lenin was moved to state, "Communism equals Soviet power plus electrification," as part of the propaganda for industrialization. Electricity took over the driving of fans, elevators, and cranes, driving coal-mining equipment, for example, and rolling mills in steel factories. The use of individual electric motors allowed astonishing advances in speed control, precision, and productivity of machine tools.

With World War II came devastation. Power stations and fuel supplies were inevitably considered as strategic targets for the bombers during the destructive aerial attacks by both the Axis powers and the Allies. By 1946, the estimated deficiency of generating capacity in Europe was 10,000 megawatts. According to anecdotal evidence, the victory bells in Paris were only able to ring out in 1945 because of electricity transmitted from Germany, where more industrial capacity of all kinds, including power stations, had survived. Whatever the truth of this may be, the security of electricity supply quickly became an issue of undisputed importance throughout Europe, and the fuels used in electrical generation were valuable resources indeed.

Coal

At the start of the twentieth century, there was a new worldwide optimism about coal as a resource that seemed to be available in almost unlimited amounts. Coal consumption levels rose steeply both in the U.S. and Europe, to reach a peak around 1914 and the outbreak of World War I. Between the world wars, consumption quantities remained almost static, particularly in the U.S., as other fuel types started to dominate the market. Reasons for this slow-down include the rising popularity of the four-stroke "Otto" cycle engine that is widely used in transportation even today as well as the commercialization of the diesel engine. These two technologies pushed fuel sources swiftly from solid to liquid fuels.

Nuclear Power

Nuclear fission was discovered in the 1930s. Considerable research occurred in those early years, particularly in the U.S., the U.K., France, Canada, and the former Soviet Union, in the design and construction of commercial nuclear

power stations. In the early 1940s, U.S. intelligence regarding Germany's promising nuclear research activities dramatically hastened the U.S. resolve to build a nuclear weapon. The Manhattan Project was established for this purpose in August 1942. In July 1945, Manhattan Project scientists tested the first nuclear device in Alamagordo, New Mexico, using plutonium produced from a uranium and graphite-pile reactor in Richland, Washington. A month later a highly enriched uranium nuclear bomb was dropped on the Japanese city of Hiroshima, and a plutonium nuclear bomb was dropped on Nagasaki, effectively ending World War II.

The nuclear power industry suffered some notable disasters during its years of technological development. In 1979, the Three Mile Island Unit 2 (TMI-2) nuclear power plant in Pennsylvania suffered damage due to mechanical or electrical failure of parts of the cooling system. Just seven years later, on the opposite side of the Iron Curtain near an obscure city on the Pripiat River in north-central Ukraine, another disaster occurred. This accident became a metaphor not only for the horror of uncontrolled nuclear power but also for the collapsing Soviet system and its disregard for the safety and welfare of workers. On April 26, 1986, the No. 4 reactor at Chernobyl exploded and released 30 to 40 times the radioactivity of the atomic bombs dropped on Hiroshima and Nagasaki. The Western world first learned of history's worst nuclear accident from Sweden where abnormal radiation levels, the result of deposits carried by prevailing winds, were registered.

Ranking as one of the greatest industrial accidents of all time, the Chernobyl disaster and its impact on the course of Soviet events can scarcely be exaggerated. No-one can predict what will finally be the exact number of human victims. Thirty-one lives were lost immediately. Hundreds of thousands of Ukrainians, Russians, and Belo Russians had to abandon entire cities and settlements within the 30 kilometer zone of extreme contamination. Estimates vary, but it is likely that over 15 years after the event, some 3 million people, more than 2 million in Belarus alone, continued to live in contaminated areas.

Oil

Often accused of being one of the two great evils in the energy sector along with nuclear power, the oil industry grew over the course of the twentieth century to acquire significance and influence previously unimagined for any industrial sector. As the century opened, the U.S. was the largest oil producer in the world, but the discovery and exploitation of reserves in the Middle East, South America, and Mexico soon shifted the balance of the market away from the U.S., which by 1950 produced less than half the world's oil. This trend continued and by the year 2000, oil production was almost equally divided between OPEC (Organization of Petroleum-Exporting Countries) and non-OPEC countries. Even in the early years of the century, the geographical spread of supply and demand quickly created the need for a system of distribution of unprecedented scale. The distances and quantities involved led to the construction of pipelines and huge ocean-going ships and tanker trucks. The capital intensive nature of these infrastructure projects, as well as the costs of exploration and exploitation of oil fields, concentrated control of resources in the hands of a few companies with vast coffers. As the reserves from easily exploitable sites dwindled, the pockets even of governments were insufficiently deep to invest in new drilling projects, and Royal Dutch Shell, Standard Oil, British Petroleum, and others were born.

Concern about fossil fuel depletion began to be voiced around the world in the 1960s, but the issue created headlines on the international political circuit in 1970 following the publication of the Club of Rome's report "*Limits to Growth*." This document warned of the impending exhaustion of the world's 550 billion barrels of oil reserves. "We could use up all of the proven reserves of oil in the entire world by the end of the next decade," said U.S. President Jimmy Carter. And although between 1970 and 1990 the world did indeed use 600 billion barrels of oil, and according to the Club of Rome reserves should have dwindled to less than zero by then, in fact, the unexploited reserves in 1990 amounted to 900 billion barrels not including tar shale.

Hydroelectric Power

Not a recent development by any stretch of the imagination, hydroelectric power was used extensively at the start of the twentieth century for mechanical work in mills and has a pedigree stretching back to ancient Egyptian times. Indeed, water power produces 24 percent of the world's electricity and supplies more than 1 billion people with power. At the end of the twentieth century, hydroelectric power plants generally ranged in size from several hundred kilowatts to many

hundreds of megawatts, but a few mammoth plants supplied up to 10,000 megawatts and electricity to millions of people. These leviathans, or "temples of modern India," as India's first prime minister Jawaharlal Nehru declared, were also the cause of massive discontent from social and environmental standpoints. The displacement of local indigenous populations and failure to deliver promised benefits were just two of the many complaints. By comparison, and despite hydroelectric power's renewable credentials, the use of conventional fossil fuel technologies such as natural gas remained relatively uncontroversial.

Coal-Gas Technology

A derivative of coal as its name implies, coal-gas is produced through the carbonization of coal and has played a not insignificant role in the development of power and energy in the twentieth century. It was an important and well-established industry product as the century opened, although electricity had already started to make inroads into some of the markets that coal-gas served. Coal-gas enjoyed widespread use in domestic heating and cooking and some industrial facilities, but despite the invention of the Welsbach Mantle in 1885, electricity soon started to dominate the lighting market. The Ruhrgebeit in Germany was the most active coal-gas producing area in the world. It was here that the Lurgi process, in which low-grade brown coal is gasified by a mixture of superheated steam and oxygen at high pressure, flourished for many years. However, as the coal supplies necessary for the process became increasingly expensive, and as oil fractions with similar properties became available, the coal-gas industry swiftly declined. In fact, when the coal industry seemed to have reached a pinnacle, another rival industry—natural gas—was being born.

Natural Gas

The American gas industry developed along different lines from the European market. Each started from a different basis at the dawn of the twentieth century. The U.S. had been quick to adopt the production of coal-gas, which was used for lighting as early as 1816. After the discovery of fields of largely compatible natural gas in relatively shallow sites when searching for oil reserves, the natural gas industry expanded swiftly. Large-scale transmission mechanisms were developed with alacrity, and one noteworthy example of this came from the Trans-Continental Gas Pipeline Corporation, which completed a link from fields in Texas and Louisiana to the demand-intensive area around New York in 1951. By contrast, in Europe the exploitation of natural gas began in earnest in the years following World War II. In the Soviet Union, for example, the rich fields around Baku in Azerbaijan were connected to both their Eastern Bloc allies by 1971 and also to West Germany and Italy by over 680,000 kilometers of pipelines.

In Western Europe developments on the geopolitical level benefited Britain, which officially acquired the mineral rights for the western section of the North Sea in 1964. Just one year later, the West Sole field was discovered. Britain had already imported some natural gas from the U.S., and within 12 years had switched almost entirely from manufactured coal-gas to natural gas. This conversion was no simple operation. The differing properties of manufactured and natural gas meant that domestic and industrial appliances numbering in the tens of millions had to be altered. The British conversion scheme, which lasted ten years, is estimated to have cost £1000 million. Other similar conversion programs were carried out in Holland, Hungary, and even in the Far East.

In October 1973, panic gripped the U.S. The crude-oil rich Middle Eastern countries had cut off exports of petroleum to Western nations as punishment for their involvement in recent Arab–Israeli conflicts. Although the oil embargo would not ordinarily have made a tremendous impact on the U.S., panicking investors and oil companies caused a gigantic surge in oil prices.

There were more oil scares throughout the next two decades. When the Shah of Iran was deposed during a revolution, petroleum exports were diminished to virtually negligible levels, causing crude oil prices to soar once again. Iraq's invasion of Kuwait in the 1990s also inflated oil prices, albeit for only a short time. These events highlighted the world's dependence on Middle Eastern oil and raised political awareness about the security of oil supplies.

The "dash for gas" in the U.K.—the rapid switch from coal to gas as the dominant source of power generation fuel—was no doubt partly instigated by the discovery of home reserves there. Worldwide the new application of an old technology, combined cycle gas turbines, or CCGTs, played a significant role. During the last decades of the twentieth century, the gas turbine emerged as the world's dominant technology for electricity generation. Gas turbine power plants thrived in countries as diverse as the U.S., Thailand, Spain, and Argentina. In the U.K., the changeover began in the late 1980s and resulted in

the closure of many coal mines and coal-fired power stations. As electricity industries were privatized and liberalized, the CCGT in particular became more and more attractive because of its low capital cost, high thermal efficiency, and relatively low environmental impact. Indeed, this technology contributed to the trend identified by Cesare Marchetti, which depicts the chronological shift of the world's sources of primary power from wood to coal to oil to gas during the last century and a half. Each of these fuels is successively richer in hydrogen and poorer in carbon than its predecessor, supporting the hypothesis that we are progressing toward a pure hydrogen economy.

Distributed Generation

Embedded or distributed generation refers to power plants that feed electricity into a local distribution network. By saving transmission and distribution losses, it is generally considered to be an environmentally and socially beneficial option compared with centralized generation. Technologies that contributed to the expansion of this mode of generation include wind turbines, which developed to the point where their cost of generation rivaled that of central power stations, photovoltaic cells, and combined heat and power units. These industries expanded massively in the latter years of the century, particularly in Europe where regulatory measures gave impetus and a degree of commercial security to the fledgling industries.

Hydrogen

Many industries worldwide began producing hydrogen, hydrogen-powered vehicles, hydrogen fuel cells, and other hydrogen products toward the end of the twentieth century. Hydrogen is intrinsically "cleaner" than any other fuel used to date because combustion of hydrogen with oxygen produces energy with only water, no greenhouse gases or particulate exhaust fumes, as a byproduct. At the close of the twentieth century, however, although prototypes and demonstration projects abounded, commercial competitiveness with conventional fuels was still only a distant prospect.

From almost wholly somatic sources of power in 1900, energy and power developed at an astonishing pace through the century. As the century closed, despite support for "green" power, particularly in developed nations, the worldwide generation of energy was still dominated by fossil fuels. Nevertheless, unprecedented changes seemed possible, driven for the first time by environmental and social concerns rather than technological possibilities or purely commercial considerations. Awareness of energy-related carbon emissions issues addressed by the Kyoto protocol raised questions concerning the institutional arrangements on both national and international levels, and their capacity for action in responding to public demand. After a century of development, a wide variety of institutional and regulatory regimes evolved around electricity supply. These most often took the form of a franchised, regulated monopoly within clearly defined administrative boundaries, in a functional symbiosis with government. However, each has the same basic technical model at its heart; Large, central generators produce AC electricity, and deliver it to consumers over a network. The continuing stable operation of this system on which many millions of people rely, once considered the responsibility of central governments, is changing. The increasing shift toward liberalization and internationalization is moving responsibility for energy supplies away from state-owned organizations, a trend compounded by the environmental and institutional implications of renewable energy technologies.

See also **Electrical Power Distribution; Electrical Energy Generation and Supply, Large Scale; Electricity Generation and the Environment; Fuel Cells; Lighting**

IAN BURDON

Further Reading

Bowers, B. *A History of Electric Light and Power*. Peregrinus, New York and Peregrinus/Science Museum, Stevenage, 1982.
Edgerton, D. *Science, Technology and the British Industrial 'Decline' ca. 1870–1970*. Cambridge University Press, Cambridge, 1996.
Marchetti, C. When will hydrogen come? *Int. J. Hydro. Energ*. 10, 4, 215–219, 1985.
Leggett, J. *The Carbon War: Global Warming at the End of the Oil Era*. Penguin, London, 1999.
MacNeill, J.R., Ed. *Something New Under the Sun: An Environmental History of the Twentieth Century*. Penguin, London, 2001.
Nye, D.E. *Electrifying America: Social Meaning of a New Technology, 1880–1940*. MIT Press, Cambridge, MA.

Engineering: Cultural, Methodological, and Definitional Issues

At the beginning of the twentieth century, engineering was engaged in transformations revolving around the nature and application of engineering knowledge and the education and social status of its practitioners. These transformations continued well

into the century, and while the pace of change had certainly abated by the end of that period, change nevertheless remained a constant throughout.

The advent of the new ‘‘science-based’’ fields of electrical and chemical engineering in the late nineteenth and early twentieth centuries lent weight to the emerging view that the relationship between science and engineering was that science discovered and explained fundamental truths while engineering applied the theories produced by science to the production of technical artifacts. This led many to divide the history of engineering into two principal periods: prescientific engineering followed by science-based engineering. The latter, of course, was seen as far more effective owing to its reliance on science.

This neat conceptualization, however, is not supported by either the historical record or actual engineering practice. Engineers throughout history up to and including the present have utilized a wide variety of knowledge ranging from rules of thumb (heuristics) to tacit understanding to graphical methods of analysis to sophisticated analytical models to scientifically derived principles of material (in its broadest sense) behavior. Various attempts at categorizing these different types of knowledge have been made (Vincenti, 1993; Addis, 1990), but they are less important for the particular organizing schemes they suggest than for their explicit observation that modern engineering is not simply or even primarily a matter of applied science (although the scientific component is certainly important). One indicator of this was Project Hindsight, a study conducted by the U.S. Department of Defense in the 1960s that painstakingly analyzed the key intellectual contributions directly underlying the development of 20 core weapons systems and revealed that the vast majority constituted technological rather than scientific knowledge.

The process by which engineering knowledge grows and advances, however, does appear to echo the process by which scientific knowledge grows and advances. In broad terms, most models of the latter still invoke the notion of paradigm shifts attributed to Thomas Kuhn (1996). This model divides scientific practice into periods of normalcy and disruption. The former is characterized by shared fundamental assumptions about the world (or at least the part of it that is relevant to the area of investigation) and the way it operates. These assumptions are so basic and ingrained that they are seldom explicitly examined. Eventually, compelling empirical evidence arises that calls into question the validity of those assumptions. A new paradigm forms to account for the new evidence (typically encountering substantial resistance) and, if it contains sufficient explanatory power, ultimately replaces the old paradigm with its new set of fundamental assumptions. Normal science then resumes within the new paradigm.

Fundamentally similar models have been proposed to describe the development of engineering knowledge. These models differentiate between normal and disruptive practice. They also tend to invoke a variation–selection–retention mechanism to explain how new solutions—embodying fundamentally new assumptions—to a problem are generated and the most effective one selected and retained.

Precisely because engineering is more than simply applied science, there is substantial room for variation in practice, including variation rooted in cultural differences. Differing professional and academic commitments to theoretical analysis on the one hand and empirical experimentation on the other, for example, will influence design goals and priorities (given that tradeoffs are inescapable) such as economy, simplicity, and esthetics. Those goals and priorities then shape both the variation and selection of technical alternatives.

While these processes for science and engineering certainly exhibit substantial similarities (in fact they represent variations on what some believe is a universal model for the generation and growth of knowledge) the values and objectives they embody are in many ways the reverse of each other. Edwin Layton Jr. (1971) has dubbed science and engineering ‘‘mirror-image twins.’’ This is most immediately obvious in their goals. Whereas science aims to understand a phenomenon, engineering aims to solve a practical problem. In the physical sciences, the more abstract and general the work the better. Specific, concrete applications tend to garner less prestige. In engineering, on the other hand, the successful design and creation of artifacts or processes usually draws the most applause while purely theoretical work is less revered. For scientists, publication of results is viewed as an integral activity while, for engineers, publication is of decidedly less importance than actual practice.

While these observations remain largely accurate, they have by no means been immutable. The professionalization of engineering and attempts to firmly ground it in science produced an impetus to publish that began in the nineteenth century and increased throughout the twentieth. This trend was reinforced by the rise of academic engineering with its inherent need for publication outlets. The latter decades of the twentieth century witnessed a partial

breakdown of the divide that had developed between practicing and academic engineers in the U.S. and the U.K. as universities were increasingly seen as a key source of new products and processes for industry. In a sense, this represented a return to the close ties between industry and engineering academics in the early decades of the century, but in this case the focal point was the production of intellectual property rather than the production of employees. This pattern was less substantial in other countries such as France and Germany, where scientific formalism in engineering had long been embraced and close cooperation between academic engineering and industry had always been prevalent.

Engineering education both motivated and reflected the move toward science-based engineering. This orientation was exactly the opposite of what characterized American engineering education through most of the nineteenth century. At the turn of the century, most practicing engineers in the U.S. had been trained through apprenticeship. Mechanical engineers were the product of a (machine) "shop culture" while civil engineers came out of a "field culture." The idea of engineers being trained in a school setting was still considered a bit odd by many, and the new scientifically oriented engineers emerging from American colleges and universities were viewed with some suspicion.

French technical education at the end of the nineteenth century, in contrast, was at its highest level the epitome of science-based engineering. French technical education was just as stratified as French society (the Revolution having dampened but not eliminated class distinctions). Its three tiers catered to very different populations and in very different ways. At the top of the hierarchy, as is the case today, were the prestigious École Polytechnique and its affiliated Écoles d'Application. The former provided education in engineering fundamentals (what today would be considered an engineering core curriculum), after which the latter would provide specialized training in a particular technical field. The vast majority of graduates ended up working for either the state or the military. The second tier consisted of the École Centrale des Arts et Manufactures, which aimed to train engineers for industry rather than government or military service. Making up the third tier were the Écoles d'Arts et Métiers, which concentrated on workshop training such as forging and machine fitting. This contrasted with the top tier, which emphasized mathematical theory. The educational thrust of the second tier was somewhere in the middle, revolving around such things as industrial chemistry and metallurgy.

Even as academia and industry throughout the industrialized world began to embrace (if they had not already) the notion of engineering as applied science, evidence of its limitations occasionally presented itself. Research at the U.S. Bureau of Public Roads between the world wars, for example, reflected the shift from empirically based research to research focused on theory and mathematical models (albeit informed by data gleaned from small-scale isolated experiments). Full-scale field studies were replaced by a search for fundamental principles that could serve as a basis for "rational" road design. This more scientific approach for road design, however, proved far less effective than the earlier efforts. Such cases reveal the complexity of the role science plays in engineering and that wholesale adoption of a scientific sensibility does not necessarily serve the ends of engineering.

Nevertheless, if science was the heir apparent to experience as a basis for engineering at the turn of the century, it was undeniably king in the aftermath of World War II. While the atomic bomb is usually seen as the most prominent example of scientific contributions to the war effort, there were plenty of others as well, including radar and the digital computer. That these were at least as much technological as scientific achievements was an unappreciated distinction. As a result, universities in the U.S. and the U.K. hastened to rid themselves of the last vestiges of practical training.

A review of university engineering education in the early 1950s sponsored by the American Society for Engineering Education fully reflected the ethos of engineering as applied science. In recognition of the increased reliance on science, the report recommended that new engineering faculty have an appropriate doctorate degree (PhD). It also called for the elimination of courses having a "high vocational and skill content" or attempting to convey "engineering art and practice" in favor of courses in engineering science, effectively sounding an official death knell for shop and field culture. The emphasis on theory and analysis was further reinforced by the general expansion of American higher education after the war, driven in part by an influx of World War II veterans. Swelling enrollments meant large class sizes, and engineering classes were no exception. Theory and analysis lent themselves to large lecture classes more readily than did design and other less scientific types of engineering knowledge.

By the end of the decade, however, employers and practicing engineers in both the U.S. and

Europe were beginning to complain of the declining ability of engineering graduates to engage in design. Accompanying these complaints was increasing criticism of engineering education that imbued students with a "blind faith" in the results of theoretical calculations and left them unable to relate mathematical engineering models to the requirements and behavior of actual artifacts. (Similar concerns have been voiced more recently regarding the results of computer-aided design tools.) Increasingly, engineering graduates, while displaying formidable analytical skills, exhibited a much-reduced ability to actually design technical artifacts. Moreover, this tendency became more pronounced with each higher academic degree. As a result, those recruited to engineering faculties were by definition those with the least inclination toward design. A 1980 international survey of engineering education found that while U.S. engineering curricula had to some extent reintroduced design as a topic of instruction, it was generally held in low esteem by the academy. In contrast, engineering education in Germany and the Netherlands incorporated a strong practical component with no diminution of status. Japan fell somewhere in between.

The twentieth century brought with it an acceleration of changes to the professional status of engineers that had been sparked by the development of large-scale industrial corporations in sectors heavily dependent on science and technology. Prior to this time, engineers, or at least those who were not part of their nation's military or government, practiced as independent professionals, typically on a contractual basis. As such, they enjoyed a degree of autonomy comparable to that of other independent professionals such as doctors and lawyers. The rise of large science and technology-based corporations changed this as, over time, increasing numbers of engineers became salaried employees rather than autonomous practitioners.

These large corporations were epitomized by firms such as General Electric (GE), American Telephone & Telegraph (AT&T) and DuPont. In addition to their need for engineers to carry out their day-to-day operations, these firms and those like them also required large pools of scientific and technical expertise to conduct research and development (R&D). These three companies, in fact, became as well known for their industrial R&D laboratories as for their other activities. That companies like GE, AT&T, and DuPont were at the forefront of this trend was not surprising. The electrical and chemical industries were considered deeply rooted in scientific knowledge right from the beginning and so were pioneers of industrial research. Other industries quickly followed suit. The years between the turn of the century and World War II saw a sweeping surge in corporate scientific and technical R&D.

This need for highly trained researchers as well as operating personnel was a key force driving the shift in engineering education from shop and field culture to a school culture. Industry and higher education worked together quite closely to shape engineering curricula that would produce employees with the requisite knowledge and skill sets. Upon entering the industrial work force, engineering graduates would often be put through internal corporate training programs designed to make them effective and loyal employees whose interests were appropriately aligned with those of their employer. Socialization was just as much an objective as technical proficiency and an understanding of company operations. The GE "test course" was one of the earliest and best known of these programs.

This shift in circumstances produced consternation on the part of engineers who worried a great deal about their status in society (and still do). This was especially true in the U.S. and the U.K., where there was a distinct absence of class-oriented mechanisms supportive of their status goals or class-based stratification that was almost wholly independent of those goals, respectively. This was unlike the situation in France, where the three-tiered system of technical education at least promised those in the top tier a modicum of professional status.

Engineers as employees confronted a fundamental tension. On the one hand, as employees they were expected to put the interests of their employers first and foremost, especially those who rose to management positions. On the other hand, as professionals they were expected to concern themselves with the interests of society as a whole. Among U.S. engineers in the early years of the twentieth century, this latter imperative crystallized under the rubric of social responsibility. Social responsibility in the sense of disinterested public service implied a measure of professional autonomy while at the same time not overly offending corporate managers.

Nowhere was this notion more firmly embraced than in the U.S. The progressive movement of the late nineteenth and early twentieth centuries had created a deep and abiding faith in the power of scientific and technical expertise in the public service. While engineers had always been involved

in the development of important infrastructure—roads, bridges, dams, and so on—they began to be perceived by many, including themselves, as essential instruments of material progress and improved quality of life. Many U.S. engineers took this perspective even further, viewing themselves as the shepherds of societal progress by virtue of their commitment to rational and impartial thought and analysis. This attitude found its fullest, albeit most futile, expression in the technocracy movement between the World Wars. In seeking to apply the methods of scientific rationalism to governance, however, technocracy seriously discredited the notion of social responsibility rather than acting as its ultimate expression.

In practical terms, this tension between professionalism and corporate capitalism frequently played itself out within the engineering professional societies. (As a result of the importance of its stratified system of technical education, sector-based professional engineering societies in France have not developed in the same way or played the same role as those in the U.S. and the U.K.) Issues of membership requirements (technical versus business), ethical codes, and disciplinary mechanisms were all areas in which the clashing priorities of engineers and managers could not be entirely avoided. These tensions were verbally reconciled by equating societal progress with technological progress, thereby making society by definition the beneficiary of corporate technical activities. When it came to actions, however, there was no escaping the fact that an insistence on professional autonomy and independent thinking at some point had to come into active conflict with the corporate ethos.

Engineering professional societies in other countries often carried with them a degree of regulatory authority for their fields, and this provided a strategic avenue that U.S. engineering professional societies lacked. In the U.K., for example, the engineering professional societies accredit curricula and nominate Chartered Engineers, a mark of technical competence and achievement. Moreover, these societies set rigorous entry requirements such that even an engineering degree from an accredited curriculum is often insufficient to fully exempt a graduate from society entrance exams. This is not to say that engineers in the U.K. or elsewhere do not worry about their status in the eyes of the public or with respect to other professions. However, certain professional structures can offer a means of at least partially addressing those concerns while others are less effective in that regard.

Engineering in the twentieth century then, is not a story of the straightforward and triumphal application of science to the creation of technical artifacts. Rather, it is a very human story of myriad motivations, perceptions, and conflicts. The engineering achievements of the century are not at all diminished by recognizing that the epistemological, educational, and professional development of engineering has been as much a social process as anything else. On the contrary, the achievements become all the more impressive, and the failures all the more understandable, with an appreciation of the nonphysical forces that have been pivotal in shaping engineering from the end of the nineteenth century to the beginning of the twenty-first.

STUART S. SHAPIRO

Further Reading

Addis, W. *Structural Engineering: The Nature of Theory and Design*. Ellis Horwood, New York, 1990.

Billington, D. *The Tower and the Bridge: The New Art of Structural Engineering*. Princeton University Press, Princeton, 1985.

Bucciarelli, L. *Designing Engineers*. MIT Press, Cambridge, MA, 1996.

Buchanan, R.A., *The Engineers: A History of the Engineering Profession in Britain, 1750–1914*. Jessica Kingsley, London, 1989.

Constant, E. *The Origins of the Turbojet Revolution*. Johns Hopkins University Press, Baltimore, 1980.

Ferguson, E. *Engineering and the Mind's Eye*. MIT Press, Cambridge, MA, 1994.

Kranakis, E. *Constructing a Bridge: An Exploration of Engineering, Culture, Design, and Research in Nineteenth-Century France and America*. MIT Press, Cambridge, MA, 1997.

Kuhn, T. *The Structure of Scientific Revolutions,* 3rd edn. University of Chicago Press, Chicago, 1996.

Layton, E. Jr. Mirror-image twins: the communities of science and technology in 19th century America. *Technol. Cult.,* 12, 1971.

Layton, E. Jr. *The Revolt of the Engineers: Social Responsibility and the American Engineering Profession*. Johns Hopkins University Press, Baltimore, 1986.

Noble, D. *America By Design: Science, Technology, and the Rise of Corporate Capitalism*. Oxford University Press, Oxford, 1979.

Petroski, H. *To Engineer is Human: The Role of Failure in Successful Design*. Vintage Books, New York, 1992.

Petroski, H. *Design Paradigms: Case Histories of Error and Judgment in Engineering*. Cambridge University Press, Cambridge, 1994.

Reynolds, T., Ed. *The Engineer in America: A Historical Anthology from Technology and Culture*. University of Chicago Press, Chicago, 1991.

Seely, B. The scientific mystique in engineering: highway research at the Bureau of Public Roads, 1918–1940. *Technol. Cult.,* 25, 1984.

Vincenti, Walter. *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*. Johns Hopkins University Press, Baltimore, 1993.

Engineering: Production and Economic Growth

Production engineering as a philosophy, a theory, and a practical technique for organizing industrial processes to optimize output and minimize waste resulted from two sources. One was the rational spirit of the eighteenth century Enlightenment, which encouraged the organization of all activities according to scientific principles. This was reinforced by a metaphysical belief that replicating in industry the rational order observed in the physical universe would reduce waste and lessen injustice to the work force. This faith encouraged several of the pioneers of management theory, including Frederick W. Taylor, F.B. Gilbreth, and the "technocrats" (see below). The other chief source of production engineering was the recurrent need from the 1850s to the present to reduce waste in the primary energy industries, and to ensure that raw materials were not being squandered. The several periods of great anxiety over fuel resources (1860s, 1890s, 1920s, 1930s, and 1970s) together with the need to organize entire nations to meet the demands of global warfare stimulated the growth of comprehensive theories of how best to manage resources, industrial production, and product distribution. The British engineer Armstrong, and the philosopher–scientist W.S. Jevons were among the first to attempt large scale reviews of energy resources, fuel consumption, and the economic consequences of production, with some attempt to foresee the consequences for future ages. Jevons great work *The Coal Question* contained most of the concepts employed by much later reviews of energy and materials use on a nationwide scale. Many pioneers of rational methods in industry were closely connected with energy use, productivity, and the physical sciences, including Louis Le Chatelier (railway mechanics; physical chemistry) and Wilhelm Ostwald (chemistry, thermodynamics). In 1887, Ostwald introduced a general philosophy in which energy was a basic concept underlying natural and industrial phenomena. Called "energism" or "energetics," this philosophy anticipated some of the later ideas of Taylor, Frederick Soddy, and the technocracy movement and proposed reforming orthodox economics to make energy rather than monetary value the basic unit for measuring production. Ostwald organized an international movement called "The Bridge" to encourage energy efficiency in all activities, with the rational spirit of science as the guiding ideal. Efficient energy use, management, and organization were advocated throughout industry, but the movement was destroyed by the onset of World War I (1914–1918).

The waste of human and material resources by nineteenth century capitalism and the destructiveness of the Great War gave widespread publicity to the writings of Thorstein Veblen who contrasted the rational, creative methods and progressive values of the engineer with the irrational, destructive methods, and commercial values of the parasitic financier, lawyer, and politician. Disciples of Veblen seized on the work being done by Taylor, Gilbreth, and H.L. Gantt as providing models not only for factories producing material goods but also for a just society that would eliminate the waste of human skills and potential. The most eager disciples of applying rational methods and scientific management on a national or global scale were the advocates of technocracy ("technocrats") in the U.S. who exerted considerable influence between 1920 and 1940. Technocracy became a political movement, with a radical agenda for social reform and reorganization of the nation along production engineering lines. The movement even had a uniform code of dress and attracted more nonengineers than engineers, although its chief exponents included the engineers H. Scott, W. Rautenstrauch, and H. Gantt. American technocrats were inclined to isolationist nationalism, but others were more internationally minded and socialist. In Great Britain, Soddy and H. G. Wells were outstanding in publicizing technocratic values. The increased use of rational methods throughout industry and business, coupled with the mass production techniques and standardization of the American motor car industry, publicized the advantages of American methods between 1890 and 1920. The terms "Taylorism" and "Fordism" (for the systems developed by Henry Ford) were loosely used to identify a philosophy for organizing engineering-based industries that drew on scientific management, work study, time and motion analysis, standardization, increased mechanization, and mass production. The techniques were condemned by conservatives as inhuman and destructive of the craft tradition in industry, though they were recognized as essential for introducing an era of high consumption and increasing productivity that would trigger an endless age of continuous economic expansion. This was done in the U.S., the first nation to make the transition from a mature industrial economy to a high-productivity, high-consumption economy circa 1920. In the U.S., the transition was made relatively rapidly, but in Great Britain, which reached maturity in the 1850s, the

change to a high-consumption economy was delayed until the 1950s due to technological conservatism. W. W. Rostow has analyzed this progression. British engineers were slow to recognize the import of Taylor's work in production engineering, and after a visit to London he described them as fixated on the form of equipment to the detriment of understanding general production theory. The Americans were quick to recognize rational management of industry as an essential part of engineering education, as were industrialized nations elsewhere such as Germany, France, and Japan; but Great Britain proved backward in this respect.

There has always been a conflict between engineers' values and economists' values. The values of the economist were generally taken to include those of the financier, lawyer, politician, and entrepreneur. The engineer visionaries, who stressed the rational, scientific nature of engineering, were a minority who exercised considerable influence through production engineering and the Technocrat movement. They received support and sympathy from a greater number of engineers who argued that short-sighted, selfish, financial considerations were delaying or even halting technological progress. Rationally planned, scientific mechanisms (which might be an industry or a new system of transport) had to be fitted into a much broader socioeconomic "receiving system" whose structure and activities were controlled by irrational pursuers of personal wealth. Exploitation of resources, including human, stopped the engineers from constructing a creative and liberating society, which would be better in the moral sense. Building on a philosophical foundation laid down in the Enlightenment, the general body of technocrats, as distinct from the political movement of that name, argued that engineering values promised better solutions to national and global problems than orthodox economics. They argued that conventional ways of assessing wealth and economic progress were irrational and inaccurate. The established engineering professions distanced themselves from this stance, which received stronger support from journalists, writers, teachers, and academics. In Britain, H.G. Wells advocated values similar to those of the technocrats, and leading engineers did support the movement, but as individuals rather than representatives of the profession. Engineers such as J.B. Henderson and A. Ewing reminded the profession that the engineer was no mere servant of money and politics but had a responsibility to a higher enlightenment. Engineers should accept responsibility for their work and the uses that others made of it. The present-day use of the term technocrat is the opposite of what was originally intended.

Production engineering was much more than production of manufactured articles with minimum waste of time, material, and energy so that better use could be made of existing plant and workforce. The need to include a widespread system of production in the analysis led to improved methods of management of personnel, training, transport of materials, and organization of subsidiary activities. The analysis passed from assessing the contribution made by a particular process and measuring the efficiency of this process within a workshop or industrial site to a general review of industrial performance and efficiency within the national or global economy. This meant finding some means of quantifying the contribution made by engineering to the economy, and relating this to a more general index of economic performance. Technocracy in its various guises stressed the scientific nature of its activities, hence measurement was essential. For many centuries, the French Wheat Price Index served as a rough guide to the fortunes of the economy of France. When the French economy achieved global significance, this index provided a general indicator. The French Wheat Price Index was chosen because there were records going back to the twelfth century. Later, the Coal Price Index served the same purpose. Attempts were made to analyze these records, using Fourier analysis, to see if the component waves could be correlated with events and so reveal what caused the fluctuations in the economy—weather, warfare, political upheaval, innovation, or discoveries. It was argued that if trend curves persist, such analysis might suggest what should be done to meet future requirements. This positivist approach, with its dangers of determinism and historicism, was used by C.O. Liljegren in 1920 to help decide future policy in marine propulsion. It gave rise to an increasingly ambitious attempt to make technological forecasting a reliable enough aid to design and planning. Two indices emerged as useful to both engineers and economists: Gross National (and Domestic) Product (GNP and GDP) and its per capita expression; and electricity consumption, usually expressed as total kilowatt hours (kWh) per year and kWh per capita. The ratio between these two indices was also judged important. After 1920, the electricity generated by an industrialized nation was used to assess its rank as a modern state. The quantity of primary fuels used to generate this

power was a measure of the efficiency of national industry. The amount of primary energy and the quantity of electricity required to generate the GNP of a nation per year was seen as an indicator of national technological and economic standing. For example, F. Quigley's study published in 1920 suggested that Britain was underelectrified and might suffer in consequence.

This analysis, much developed and refined, enjoyed widespread use during periods of energy crisis, and it remains in general use.

After 1920, Taylorism, Fordism, and technocracy came together to create an engineering-based approach to global production, resource use, and economics. This intensified the clash with orthodox economics, politics, and finance, and widened the gap between the Technocrats and the majority of conservative, professional engineers who feared involvement in radical politics. The Depression that began in the U.S. in 1929 and spread worldwide provided the technocrats with an opportunity and, in North America, gave the Technocrat movement its most influential period. In the USSR, Germany, Italy, Japan, and other industrial countries dominated by totalitarian regimes, technocracy came to mean the use of engineering to serve a military dictatorship with maximum efficiency and minimum considerations of conscience. As a result, the word technocrat came to mean an obedient expert who discharged assigned duties with technical competence in the service of the powerful; it has never recovered its original meaning. In the 1930s the original ideals of technocracy were pursued largely in North America and Great Britain. The technocrats in the U.S., despite being isolationist and nationalist, were the most influential inside and outside North America. The technocrats argued that the failure of the financial system in a country full of skilled workers, competent engineers, and up-to-date factories was proof that the old economics should be scrapped. The U.S. workforce wanted to work. The equipment was there. The energy was there. The engineering intelligence was there. But the financial system could not facilitate turning these resources into productive activity. Leading technocrats such as H. Scott were inspired by H.G. Wells and F. Soddy and advocated making energy units the basic currency in a new economics. Echoing Ostwald, Soddy and others, Scott said that all goods and services were converted energy, and a scientific review of a nation's activities meant quantifying all human, natural, and machine activities in energy units, which could then be used as the price of goods and services. The matter of thermodynamic energy grade seems to have been sidestepped. Between 1932 and 1933, Scott, Rautenstrauch, and Hubbert compiled an energy survey of North America which gave widespread currency to many of the techniques, concepts, and terminology still found in energy analysis. Hubbert's contribution was outstanding. The survey charted the growth of 3000 industrial and agricultural products between 1830 and 1930 and measured production in terms of energy expended, volume of production, rate of growth, manpower per unit of production, power per unit of production, total power, total number of employees, and production man-hours. It was a standards-setting exercise, taken up by industrial nations and now a regular technique whose findings can be found in the annual volumes of statistics issued by governments all over the world. Scott's theory of making energy into a currency was successfully resisted by orthodox economists who used errors in the energy survey to discredit the ideology and political program of the technocrats. Their political program was outflanked by Franklin Delano Roosevelt's New Deal, but their analysis of energy and material use was greatly developed and widely applied during World War II and afterward during the Cold War and the energy crisis of the 1970s. The major protagonists in the World War and the Cold War embraced Taylorism and Fordism to various degrees, helped or hindered by political and ideological factors. Exploring the link between energy flows and money flows in the economy was continued and enjoyed considerable vogue in the late 1970s following the 1973 energy crisis. Despite quantification of energy investment in most goods and services, no national economy was placed on an energy-value basis. Orthodox economics and financial methods continued to dominate worldwide.

The rise of technocracy coincided with attempts to develop econometric analysis of engineering change and its consequences for the economy. Many technocrats employed econometrics, but not all econometricians supported technocracy. Christiaan Huygens, Christopher Polhem, Napoleon, Armstrong, and Jevons were a few of those who recognized the importance of engineering to a nation's economy between 1600 and 1900, but they lacked a comprehensive model of economic growth related to history. Between the world wars several comprehensive models were put forward. These models assumed that the global economy was dominated by a relatively small number of leader nations such as France and Britain in the eighteenth and nineteenth centuries

and the U.S. and Japan in the late twentieth century. In 1925 N.D. Kondratieff argued that analysis of economic performance during the industrial period showed that it could be divided up into successive cycles of growth, prosperity, stagnation, recession, recovery, and so on. He further claimed that these phases repeated themselves at regular intervals of about 53 years. As long as the structure of the model lacked a rational explanation and the precise periodicity was claimed, Kondratieff's cycles were regarded as belonging to speculative metaphysics. During and after World War II, however, a school of econometrics developed that accepted the cycles as established by reliable data and looked for an explanation. S. Kuznets and Joseph A. Schumpeter argued that global economic history was associated with distinct phases that were caused by technological innovation. Innovation created new industries, and a relatively small number of industries cross-fertilizing each other launched a new era of economic development. The industries that dominated the succeeding phase of economic growth were strategic industries, created by strategic innovations. They fostered new standards of workers' skills and new management techniques, raised standards of required engineering science, and generated a fresh understanding of what the contemporary age meant by modern technology. In their periods of rapid growth, these industries were very profitable and attracted much investment. G. Mensch argued that as they became established and less modern, these industries became less profitable and attractive to investors, although they could still be important in the economy. They might be profitable enough to make the investments in them worth maintaining, but eventually the diminishing returns on investment would encourage the creation of new, dynamic industries made possible by the most recent generation of technological innovations.

Supporters of this theory claimed evidence from history. The mechanized-industrial age was launched in Great Britain by industries based on coal, iron, and textiles with transport by canal. These began in relatively few centers (Ironbridge, Cromford, and Manchester). The skills cultivated at all levels in the "first industrial nation" could not be learned easily or quickly in other nations, and Britain enjoyed a practically unassailable lead. The strategic innovations that created the industries that dominated the next phase were steam power, application of steam power to older industries, and railways. These grew out of the older technologies created in England, and so Britain enjoyed prolonged leadership in the global economy. Later, periods were dominated by new technologies created relatively rapidly and did not evolve out of older industries. During such change points, leadership in economic growth passed to nations that developed the new skills and cultural values crucial during the next phase. Examples quoted were Germany and the U.S. in the period after 1890 when electrical engineering, industrial chemicals, and the automotive industry were strategic. Later still, electronics, aviation, rocketry, nuclear engineering, computing, and the technologies of the post-1945 age ushered in new phases of development and witnessed the rise of Japan. Though some econometricians accepted the precise periodicity of the Kondratieff model (as did Mensch), a larger number accepted that the interpretation of industrial growth was roughly correct and could be used as a guide. Many were skeptical and regarded the lessons derived from the "long-wave analysis" as due to hindsight, although the classifications might be beneficial to historians. Philosophers were suspicious of implicit determinism and historicism in the models. Much criticism was directed against the creation of long-wave trend curves, or continuous traces obtained by plotting an index of economic performance against a measure of input to industry, or against time (date). How could a major innovation be associated with a particular date? Was it legitimate to create a model by treating a succession of innovations as if each were equal in economic importance to the rest? Mensch's work attempted to deal with this issue but continued to attract adverse criticism.

The use of trend curves played a major role in large-scale econometrics. Trend curves were also used on a smaller scale to judge the extent to which a particular technology, design, or product was worthy of further investment or was approaching obsolescence. If recognized as near obsolescence, it could be replaced by a successor introduced in an orderly manner, which reduced the waste of unused potential in the old technology. The works of B. Twiss and R. Foster illustrate the use of S-curve analysis in management and business circles. Attempts to integrate small- and medium-scale S-curve analysis with the large-scale, long-term models of Kuznets, Schumpeter, and Mensch have not yet succeeded. These theories are often used as primarily qualitative guides based on history. As such, they provide valuable lessons. They suggest that it is destructive of a nation's standing, or an industry's profitability, if the nation or industry maintains investment (including intellectual skills) in declining activities that were once

strategic and fails to reorganize to take control of completely new strategic industries based on recent innovations. Industrial and economic leadership may be associated in future with global networks rather than individual companies located in one nation. The changing nature of engineering is leading to industries based on nanotechnology, artificial intelligence, cybernetics, and biotechnical hybrids. The meaning of "industry" and "product" is being redefined. Contemporary and future industries may increasingly produce knowledge, patent rights, and licenses to manufacture as earlier industries produced steel and heavy equipment. The manufacture of older technologies is shifting to industrial cultures outside the first rank. This shift has caused widespread reorganization of engineering education and the profession in older industrial countries such as the U.K.

The 1973 energy crisis and the ongoing discussion concerning sustained growth, limits to growth, and environmental damage due to industrial activity gave a great impetus to neotechnocracy, long-wave econometrics, and engineer values. A few of the original technocrats still lived and carried on their original campaign, though technocracy, which survived as a movement, enjoyed little influence. The growth of innovation analysis and the need to assess the worth of expensive military projects revived interest in technological forecasting, which received funds from military sources and other government departments. The anticipated shortage of fossil fuels in the 1970s and 1980s focused attention on using trend curves to assess nearness to exhaustion of resources. Use was made of similar trend curves to link industrial production to damage to the environment. The link between energy consumption per head and GNP per capita was calculated for different countries at various stages of industrial development and used to calculate how much fossil fuel and raw materials would be needed to raise the poorer nations to the standard of living in the U.S. or Germany. Though the link between GNP per capita and energy consumption per capita was condemned as misleading, the analysis indicated the probable impossibility of abolishing world poverty in this sense, taking into account population growth, lengthening life spans, annual expansion of the economy, and expectations of a regular increase in standard of living. During the 1970s, the Florida School of analysts, associated with H. T. Odum and E.C. Odum produced studies of energy flow correlated with money flow in society and linked money value to energy value, along lines similar to those pursued by the various technocrats in earlier years. The Odums concluded that whereas lack of money in circulation was the problem in 1929, the cause of postwar economic crises was more likely to be limited access to cheap energy and raw materials. In the 1970s the aggressive "production engineer values" of the period from 1910 to 1940 were less in evidence. The "engineer values" and the technocracy were still there, but they were presented in a more circumspect manner and in closer association with a liberal, enlightened economics that accepted limits to growth on environmental grounds.

Development of these theories and philosophies continues, as does the clash between engineers' values and economists' values. Some states require that any energy-consuming scheme be subjected to an analysis beforehand to calculate the total energy and resources investment in the project compared with the anticipated benefits. Many nations now calculate, as part of the GNP assessment, the energy investment in the goods produced and services provided.

See also **Engineering: Cultural, Methodological, and Definitional Issues.**

Michael C. Duffy

Further Reading

Akin, W.E. *Technocracy and the American Dream: The Technocrat Movement 1900–1941*. University of California Press, 1977.
Ewing, A. An Engineer's Outlook. Annual Report. British Association for Advancement of Science, 1931, pp 1–19.
Foster, R. Innovation: *The Attacker's Advantage*. Macmillan, London, 1986.
Fox, R., Ed. *Technological Change*. Harwood, 1996.
Gilbreth, F.B. *Motion Study*. Van Nostrand, New York, 1911.
Gilbreth, F.B. *Primer of Scientific Management*. Van Nostrand, New York, 1912.
Grubler, A. *Technology and Global Change*. Cambridge University Press, Cambridge, 1998.
Henderson, J.B. Invention as a Link in Scientific and Economic Progress. Annual Report. British Association for Advancement of Science, Leeds, 1927, pp 120–137.
Jevons, W.S. *The Coal Question: An Inquiry into the Progress of the Nation, and the Probable Exhaustion of Our Coal Mines*, reprint of 3rd edn. Kelley, New York, 1965.
Kondratieff, N.D. The Major Economic Cycles. Translation of Russian 1925 edn., *Review*, 4, pp 519–562, 1979.
Kuznets, S. *Economic Change*. W.W. Norton, New York, 1953.
Landes, D.S. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press, Cambridge, 1969.

Liljegren, C.O. Coal, Oil or Wind? *Trans. Inst. Eng. Shipbuilders Scot.,* 64, 242–302, 1920–21.
Mensch, G. *Stalemate in Technology: Innovations Overcome the Depression*. Ballinger, Cambridge, MA, 1979.
Odum, H.T. and Odum, E.C. *Energy Basis for Man and Nature*. McGraw-Hill, New York, 1976.
Pavitt, K., Ed. *Technical Innovation and British Economic Performance*. Macmillan, 1980.
Quigley, H. *Electrical Power & National Progress*. George Allen & Unwin, 1920.
Rostow, W.W. *The Stages of Economic Growth,* 2nd edn. Cambridge University Press, Cambridge, 1971.
Schumpeter, J. *Business Cycles: A Theoretical, Historical & Statistical Analysis of the Capitalist Process*. McGraw-Hill, New York, 1939.
Schumpeter, J. *Konjunkturzyklen*. Gottingen, 1961.
Soddy, F. *Wealth, Virtual Wealth and Debt: The Solution of the Economic Paradox*. E.P. Dutton, New York, 1926.
Taylor, F.W. *The Principles of Scientific Management*. Harper Bros., New York, 1919.
Twiss, B. *Managing Technological Innovation,* 2nd edn. Longman, London, 1980.
Urwick, L. and Breck, E.F. *The Making of Scientific Management,* 2 vols. Pitman, 1951, 1953.
Walker, M. The Call to the Engineer and Scientist. Annual Report: British Association for Advancement of Science (York Meeting), 1932, pp 131–146.
Veblen, T. *Theory of the Leisure Class*. Macmillan, New York, 1899.
Veblen, T. *The Engineers and the Price System*. Huebsch, 1921.
Veblen, T. *The Instinct of Workmanship,* Macmillan, 1914. Reprinted by W.W. Norton, New York, 1964.

Entertainment in the Home

When Buckminster Fuller's futuristic Dymaxion House was displayed in Chicago in 1929, the television set, phonograph, and radio it featured were far from the everyday reality of most of the people who saw it. Less than 25 years later, these devices had become commonplace in middle-class homes across North America and Western Europe. The most significant forms of home entertainment technology of the twentieth century can broadly be divided into the audio and the visual, screen and sound, though as the century progressed the two categories increasingly overlapped. Electrification was a critical factor in the development and diffusion of these new entertainment technologies, as it was for other domestic technologies that became popular in the postwar period. Although some electronic devices could be operated by means of batteries, the consistent, popular use of most entertainment technologies in the home would not have occurred without mass domestic electrification. In the U.S. this occurred (unevenly) in the second decade of the century, and in Britain by 1926. Before World War II, live and recorded music and screen entertainment was more easily accessible to most people in public venues such as music and dance halls, cafes, and movie theaters. To some extent, the new technologies encouraged the return of public forms of entertainment to the domestic sphere.

In America, the phonograph displaced other forms of home music making, the most prominent of which was the piano in middle-class homes. By 1900, the Victor Talking Machine Company was marketing a domestic phonograph, and a few years later the Victrola became the first mass-marketed, enclosed phonograph. Its name was appropriated for the generic phonograph designed as household furniture. Futuristic designs eventually supplanted traditional disguises, and the phonograph was combined with radio and with audio cassette functions (in the early 1970s) in home stereo systems. By 1988, compact disk (CD) sales had surpassed the sales of phonographic long-playing albums, or LPs. Digital technology, which surpassed the phonograph record in quality of sound reproduction, durability, and convenience, was rapidly displacing the phonograph as the primary form of home sound technology.

The wireless technology that supported radio was a product of the late nineteenth century, but it did not become popular in homes until the 1920s. Prior to this, radio was the domain of industry, national security, and private enthusiasts. The simplification of technical controls to two knobs, the replacement of headphones with a loudspeaker, and the physical displacement of the radio from the basement and garage to the living room, transformed a primarily male activity into a genteel, feminized, domestic amusement for the whole family. By 1950, approximately 95 percent of American households had radios. In the 1990s, Canada pioneered digital radio, and in 1995 the first digital radio receivers were marketed to consumers. Researchers have predicted the full transition from AM–FM to digital radio by the first decades of the twenty-first century.

Television (TV), a more expensive and complex technology, was nevertheless much more quickly adopted by consumers than radio. It was introduced to the public in 1924 at Selfridges department store in London, England, by Scotsman John Logie Baird, but only became a household commodity after World War II. Much as inside and outside was merged in modern houses by means of the picture window, the introduction of television brought public culture into private homes. RCA and DuMont offered the first sets (all with black and white pictures) to the public in 1946. Between 1948 and 1955, nearly two thirds of

American households installed television sets, and by 1960 almost 90 percent of households had at least one television. The way people watched television during the last half of the twentieth century changed considerably due to the development of television-related technologies such as coaxial cable, communications satellites, the video cassette recorder (VCR), and the remote control for all of it. Cable, which spread slowly from the 1950s through the 1970s, and communications satellites, which gained popularity in the last two decades of the century, broadened the range of programming available to viewers. The digitization of television in the form of high-definition television (HDTV) began at the end of the twentieth century but was not expected to have a large-scale impact in homes until well into the first decade of the next century.

The VCR was the first peripheral device to television to gain widespread consumer acceptance. It had two distinct technological functions that together performed the important cultural task of giving consumers the freedom to choose programming outside of regular broadcasting. One function was time shifting, the ability to record a program for playback at a later time. The second closely related function allowed the replay of prerecorded cassettes, either commercially produced or made personally with home video recorders, introduced by Sony in 1980 with the Camcorder. In 1956, the Ampex Corporation of California introduced the industrial precursor to the consumer VCR: the VTR, or video tape recorder. By 1966, Ampex had sold over 500 VTRs for home use, but with open-reel video tape these were both unwieldy and expensive. In 1972, Sony introduced the U-matic, which sold for $1600 and was the first video cassette recorder specifically intended for home use. Although Sony's Betamax format video cassette was widely popular among consumers, the Video Home System (VHS) from JVC became the standard format until the last years of the twentieth century when laser disk technology began to make significant inroads. By 1988, more than half of television-owning households also had VCRs. In 1996, the first digital video/versatile technology (DVD) hit the consumer market in Japan and reached the U.S. market the next year. By the end of the century, DVD technology had become mainstream, and home video providers began to stock video disks alongside video tapes. Unlike video tapes, DVDs could be played independently of TV on home computers.

Home video games are considered television peripherals since they rely almost exclusively on the television screen to express their video component. The first generation of home video games was ushered in with the Magnavox Odyssey in 1972. Early versions were limited to preprogrammed games, whereas the popular Sony, Sega, and Nintendo systems developed in Japan offered the possibility of programmability and an endless number of games along with advanced graphics and fuller sound. Products such as Nintendo's Game Boy and the Sony Playstation were the objects of a home entertainment craze; over 10 million Playstations were sold between 1995 and the first years of the twenty-first century.

The practices of watching television, listening to music, and operating computers influenced the way people thought about, used, and designed their homes: from living rooms, to the recreation room/family room/games room, the kitchen, the home office, and even the bedroom and bathroom. Many postwar television watchers designated a TV room that set apart television watching architecturally from other domestic activities. Others had a TV area in the living room, or swivel stands that allowed for flexible viewing arrangements. Home entertainment technology also inspired the design of new types of furniture that reflected both the utopian and nonutopian expectations with which modernity and its products were received. Designs for cabinets and stands for televisions and stereos highlighted or hid their conspicuous futuristic designs. Alternatively, designers attempted to integrate new technologies into more traditional domestic environments by designing decorative cabinetry in fine woods to match preexisting traditional furniture.

See also **Audio Recording (all); Radio: AM, FM, Analogue, Digital; Radio, Early Transmissions; Television, Cable and Satellite; Television, Digital and High Definition; Television Recording, Tape**

JENNIFER COUSINEAU

Further Reading

Braden, D.R. *Leisure and Entertainment in America*. The Museum, Dearborn, MI, 1988.

Dale, R. and Weaver, R. *Home Entertainment: Discoveries and Inventions*. Oxford University Press, Oxford, 1993.

Ellis, D. *Split Screen: Home Entertainment and the New Technologies*. Friends of Canadian Broadcasting, Toronto, 1992.

Forty, A. *Objects of Desire: Design and Society from Wedgwood to IBM*. Pantheon, New York, 1986.

Herz, J.C. *Joystick Nation*. Abacus, London, 1997.

Peiss, C. *Cheap Amusements: Working Women and Leisure at Turn of the Century New York*. Temple University Press, Philadelphia, 1986.

Read, O. and Welch, W.L. *From Tin Foil to Stereo: Evolution of the Phonograph,* 2nd edn. SAMS, Indianapolis, 1976.
Spiegel, L. *Make Room for TV: Television and the Family Ideal in Postwar America*. University of Chicago Press, Chicago, 1992.
Spiegel, L. *Welcome to the Dream House: Popular Media and Postwar Suburbs*. Duke University Press, Durham, NC, 2001.
Wright, G. *Building the Dream: A Social History of Housing in America*. The MIT Press, Cambridge, MA, 1992.

Useful Websites

Schoenherr, S. Recording Technology History: http://ac.acusd.edu/history/recording/notes.html

Environment and Electricity Generation, *see* **Electricity Generation and the Environment**

Environmental Monitoring

As a dynamic system, the environment is changing continually, with feedback from both natural (climatic or biogeochemical) and anthropogenic (human activities) sources. Assessing the rate and magnitude of environmental processes is difficult, especially as data collection over time is limited to the last century, or even the last few decades. Since the 1980s, environmental monitoring programs have been developed as a response to concerns that environmental impact or sustainability of policy initiatives could not be evaluated adequately. So-called *State of the Environment Reports* date from this period. Many are concerned with the state of national, regional, or local environments (land, rivers, or seas); others focus on environments at particular risk, mostly due to human impact and pollution (e.g., environmental contaminants in marine or terrestrial wildlife), or those where environmental quality is significant in the context of human health (e.g., urban air quality; water resources, fisheries). Many monitoring programs include information collected remotely by satellites (see Satellites, Environmental Sensing), but this entry focuses on technologies and policies for *in situ* monitoring.

Adequate and sustained monitoring and its evaluation provides early warning of possible environmental degradation. Such information is important for the prediction of change that may be generated in the wake of a development project, such as dam construction or deforestation programs. In this context—as an element of an environmental impact assessment—monitored data are valuable for an evaluation of sustainability.

Monitoring involves the regular or continuous collection of specific parameters that depends on the type of environment being monitored, and can contribute to the identification of cause and effect relationships. Physical, chemical, and biological data may be collected in subsystems within specific environments. For example, certain organisms may be sensitive to changes in pH or nutrient status in aquatic environments, and so alterations in their population numbers and population composition can be used to detect change in water quality. These are called indicator species. Thus biological components are monitored to detect change in the chemistry of rivers and lakes. A range of socioeconomic data may also be monitored; factors such as population change and a drive to increase agricultural production because of the emergence of new markets are driving forces in environmental change and are likely to stimulate alterations in the physical environment.

There are three basic types of monitoring—baseline, impact, and compliance—that reflect different stages of development. Baseline monitoring involves collection of basic environmental data (e.g., soil and water pH) prior to the start of a development project. Impact monitoring focuses on the collection of both environmental and, where relevant, socioeconomic data, during the construction phase of the project. Compliance monitoring requires regular, often periodic, data collection to determine environmental quality and change after project completion and is a means of assessing if standards of environmental protection are adequate and in accord with environmental law. Through the provision of environmental data and its evaluation, monitoring programs also contribute to the enhancement of public understanding of local and regional environments and the operation of people–environment relationships.

No monitoring program can be truly comprehensive because the physical and human environments comprise a vast range of factors that interact in complex ways. Thus an initial and vital prerequisite for any monitoring program is the choice of appropriate variables that, in turn, are dependent on the objective of the project and the goal of the monitoring program. Broadly, such variables can be classified as state, pressure, and response indicators. State indicators reflect the condition of the environmental system; examples include the composition and density of the vegetation cover, soil pH and nutrient status, and water pH and chemistry. Pressure indicators may be direct measures; for example, rate of land clearance, length of fallow, agricultural productivity, rate of wood

removal from forests, rate of waste production; or indirect measures; for example, calorie consumption per individual, volume of crops produced per unit area, volumes of crops, meat or wood exported, rate of water consumption, income *per capita*, and rate of population growth. Response indicators may include soil and water conservation measures, land abandonment, and emigration rates. All may be qualitative or quantitative, or both. Each indicator and its measurement must be scientifically sound and sensitive to change. They should not cause problems for local people but should involve them where possible, as local knowledge is all-important. Each indicator and measurement must be reproducible and feasible in terms of scale, time, and finance.

Monitoring may be undertaken by various organizations, both government-related and nongovernmental. For example, the European Environment Agency of the European Union produces yearly environmental indicator reports entitled *Environmental Signals;* a special report for the millennium was also produced (OOPEC, 2002). In the US, the Environmental Protection Agency is one of several government bodies involved with environmental monitoring; between 1996 and 2001 it operated a program entitled Environmental Monitoring for Public Access and Community Tracking (EMPACT), which involved a wide range of projects on air, water, and soil monitoring in rural and urban environments. Nongovernmental agencies that produce reports on the state of the global environment based on national statistics include the Worldwatch Institute in Washington D.C.; its annual publication for the last 12 years is entitled *Vital Signs*. This publication collates statistics on "key indicators," including atmospheric, energy, agricultural, food, trade, population, disease, communication, transport, and military trends. The Worldwide Fund for Nature (WWF) produced two reports entitled *Living Planet*, in 1998 and 2002, which detail the state of the world's ecosystems and the impact of humanity on them; the data are presented on a country-by-country basis and are used to derive a Living Planet Index and Ecological Footprint which can be used for intercountry and regional comparisons. The World Conservation and Monitoring Centre, initially established in 1988 and formally linked with the United Nations Environment Programme (UNEP) since 2000, has a more specialized role insofar as it collects, collates, and evaluates worldwide data on plants, animals, and microorganisms and the ecosystems to which they belong.

Detection and characterization of environmental pollutants may be by continuous source monitoring, for example of stack emissions, industrial wastewater, or leachate sampling from landfills, or by sampling at fixed points within an areal network. The latter includes sampling of atmospheric aerosols (airborne solid or liquid particles, such as nitrous oxides and diesel particulates) at urban sites, or groundwater monitoring for nutrients and pollutants such as volatile organic compounds (benzene), heavy metals (lead, mercury, arsenic), pesticides (DDT, atrazine), polychlorinated biphenyls (PCBs), and pathogens from animal or human waste.

Indoor and outdoor ambient air quality is often measured by drawing air into an evacuated, contaminant-free container or pumping a sample through an impregnated filter paper or gas sample tube. Some optical instruments measure particle size and the amount of airborne dust particles. Colorimetric indicator methods, in which a color change of the sorbent in a gas tube gives presence or absence information, are specific to a single chemical such as arsenic.

The development of analytical chemistry techniques for chemical process control provided ready-made solutions for environmental monitoring. Infrared spectrometers, developed in the 1920s as a laboratory analytical tool, became less bulky from the 1950s and with the advent of powerful microprocessors became routinely used for measuring gaseous air pollutants, from ambient air, combustion sources, and toxic waste incinerators. Gas chromatography, a versatile analytical technique developed in the 1950s, is now available as portable units that can be used on-site for real-time direct and continuous measurement, with data logged or transmitted via radio telemetry if in a remote area. Mass spectrometry can give very fast positive identification of a broad range of compounds, whether gases, liquids, or solids. Gas-sensitive solid-state sensors, first researched in the late 1950s and commercially available from the 1970s, are now in widespread use.

Neutron activation analysis, a laboratory technique in which a sample is irradiated with high-energy neutrons, permits measurements of trace elements in atmospheric, soil and water pollution studies. Detection limits are in the parts per million to parts per billion range.

See also **Satellites, Environmental Sensing; Technology, Society and Environment**

A. M. Mannion

Further Reading

Mannion, A.M. *Dynamic World: Land-Cover and Land-Use Change*. Arnold, London, 2002.
OOPEC. *Environmental Signals 2002: Benchmarking the Millennium*. European Environment Agency Regular Indicator Report, OOPEC, Luxembourg, 2002.
Worldwatch Institute. *Vital Signs 200*. Worldwatch Institute, Washington D.C., 2003.
Worldwide Fund for Nature. *Living Planet Report 1998*. Worldwide Fund for Nature, Gland, Switzerland; New Economics Foundation, London; World Conservation Monitoring Centre, Cambridge, 1998.
Worldwide Fund for Nature. *Living Planet Report 2002*. Worldwide Fund for Nature, Gland, Switzerland, 2002.

Useful Websites

Environmental Protection Agency (EPA), Environmental Monitoring for Public Access and Community Tracking (EMPACT): http://www.epa.gov/empact/
Environmental Protection Agency (EPA), Environmental Technology Verification Program: http://www.epa.gov/etv/
World Conservation and Monitoring Centre (WCMC): http://www.unep-wcmc.org

Error Checking and Correction

In telecommunications, whether transmission of data or voice signals is over copper, fiber-optic, or wireless links, information coded in the signal transmitted must be decoded by the receiver from a background of noise. Signal errors can be introduced, for example from physical defects in the transmission medium (semiconductor crystal defects, dust or scratches on magnetic memory, bubbles in optical fibers), from electromagnetic interference (natural or manmade) or cosmic rays, or from cross-talk (unwanted coupling) between channels. In digital signal transmission, data is transmitted as "bits" (ones or zeros, corresponding to on or off in electronic circuits). Random bit errors occur singly and in no relation to each other. Burst error is a large, sustained error or loss of data, perhaps caused by transmission problems in the connecting cables, or sudden noise. Analog to digital conversion can also introduce sampling errors.

Early Error Detection

Claude Chappe's semaphore visual telegraph system in late-eighteenth century France incorporated rudimentary error checking: an operator had to check that the next station was correctly reproducing the signal. Error correction also existed in the form of a single control signal that signified "erase the last signal sent." The correct signal was then repeated.

In the early twentieth century, various techniques were applied to telegraph signals in order to prevent or reduce errors from transmitted signals. For cabled telephony, low transmission rates, shielding (to reduce cross-talk), and repeaters (to overcome signal loss due to attenuation) all helped to prevent errors. Errors could also be reduced by sending each message multiple times.

Telex over radio (or TOR), developed in the 1960s for sending telex to ships, converted binary code (usually the 5-bit Baudot code with 5 bits representing each text character) received by radio to text. In wireless transmissions, noise and fading cause missed or incorrect characters. Two similar and simple error detection and correction systems were developed: sending each character twice (the receiving station would request a resend if it did not see each character twice in sequence), and a read–confirm system, where each character was confirmed as received.

From the 1960s ASCII (American Standard Code for Information Interchange) replaced the Baudot code in teletype transmission: ASCII used seven bits to represent each text character, but an eighth bit called a parity bit could be added to each character to do an error check. Developed for teletype, ASCII also provided a common computer code used in most computer operating systems, e-mails, and HTML documents.

Parity checking is the oldest and simplest example of error checking code, used in teletype and computer data storage. A single bit, a parity check bit, is added to the transmitted data. The sum of a fixed number of bits can be made even (or odd) by properly setting the parity bit to a one or zero. Errors are detected on the receiving end simply by checking whether each received word is even (or odd). As with TOR, the receiver could request a retransmission of the message if the message and its error checking code are in conflict (or if the code is not received.) Check digits are similarly used embedded in credit card numbers and ISBN numbers to detect mistakes.

Parity checking has a low error detection rate, only about 50 percent. In block checking (also called longitudinal redundancy checking), a block check character (BCC) is added to the end of each block of data (a group of bits transmitted as a unit) before transmittal: the block check character is computed using binary addition to maintain parity counting longitudinally through the block. The checksum and cyclic redundancy check (CRC) algorithms similarly compute block check charac-

ters for each message block, but the character (for checksum) or series of characters (for CRC) added is a mathematical function of all the data in the message. As with parity checking, the receiver computes a second block check character to compare with the received character to determine whether the transmission is error free. Redundancy checking, which can be easily implemented and decoded in hardware, can detect up to 99 percent of errors, and is useful where it is easy to retransmit data. In TCP (the Internet's Transmission Control Protocol) checksums are computed for each message packet to detect transmission errors, and the receiver acknowledges only packets where the checksum is correct.

Retransmission is the oldest form of error correction—simply, the message is retransmitted until it is received without error. This is often called Automatic Repeat Request (ARQ) and is used in modems and increasingly in mobile Internet via wireless cellular networks. In stop-and-wait ARQ the sender waits for a reply in a fixed time frame after sending a signal. If there is no reply, the block is resent. In continuous ARQ, the receiver automatically starts a request for retransmission when an error is detected.

ARQ and parity checks are backward error correction (BEC) techniques used widely in computer communications; "backwards" because the sender is responsible for retransmitting any data found to be in error. In addition to error-detecting codes there are error-correcting codes—referred to as forward error correction (FEC), since the receiver corrects the error without requiring further input from the sender.

Error Correction Coding

Late-twentieth century error coding techniques are based on information coding theory, from work by Claude Shannon in 1948 (see Information Theory). "Information" is mathematically quantifiable as binary codes, as described above. Error correction coding adds extra bits (the "code word") to the data before transmission, which the receiver can decode to check if errors occurred during transmission. Simply repeating each bit a set number of times and assuming the most frequent value is correct can automatically correct for errors, but high repetition rates are not desirable. Shannon's coding theory showed that it was possible to encode messages in such a way that the number of extra bits transmitted was as small as possible (the Shannon limit), though his theory did not produce any actual ideal coding algorithms.

The first practical error correction codes were developed in 1948–1950 by Richard W. Hamming, a colleague of Claude Shannon's at Bell Labs, assisted by Marcel J.E. Golay. Hamming was working with an early relay computer, and was frustrated by having to restart his computations when the computer detected errors. Hamming and Golay "block codes" are linear algebraic rules for converting an information sequence of bits, of length k, say, into a transmitted sequence of length n bits, and had information transmission rates more efficient than simple repetition.

The Hamming and Golay correcting block codes are easy to encode and decode. In 1960 Irving S. Reed and Gustave Solomon (researchers at the Massachusetts Iinstitute of Technology) devised further error-correction block codes. More recently, convolution codes—in which each encoded block depends not only on the corresponding k-bit message block at the same time unit, but also on m previous blocks—have been used.

Error coding has great economic value and is used in many digital applications such as computer memory, magnetic and optical data storage media, radio-frequency links for space and satellite communications systems, computer networks, and cellular telephone networks (called channel coding).

See also **Electronic Communications; Information Theory; Telephony, Digital**

GILLIAN LINDSEY

Further Reading

Costello, D.J. Jr.; Hagenauer, J., Imai, H., Wicker, S.B. Applications of error-control coding. *IEEE Trans. of Inf. Theor.*, 44, 6, 2531–2560, 1998.

Berlekamp, R. *Key Papers in the Development of Coding Theory*. Institute of Electrical and Electronics Engineers, New York, 1974.

Hill, R. *A First Course in Coding Theory*. Oxford University Press, Oxford, 1986.

Nebeker, F. *Signal Processing: The Emergence of a Discipline, 1948 to 1998*. Institute of Electrical and Electronics Engineers History Center, 1998.

Experimental Stress Analysis

This branch of technology deals with the means of measuring strains in materials under load and, from these strains, inferring the stresses actually endured by the material. The fundamental idea underlying the design of all components of structures and machines that must carry loads is that

the stress in the material should be less than, or equal to, a certain prescribed level.

Unfortunately, it is not possible to measure stress directly. Stress values inside a material must be calculated using mathematical models of both the structure and properties of the material of which it is made. When fundamental material properties such as strength and stiffness (Young's modulus) are experimentally tested, the structure is kept very simple—a wire for tests in tension or a supported beam for tests in bending. Measurements of load and the extension or deflection of these structures are then used to calculate internal stresses for simple tension or compression and for simple bending theory in the case of beams. Modern high-speed computers have enabled more complex mathematical models and have rendered complicated structures amenable to theory.

While it was common from the early nineteenth century to test full-size and scale-model prototypes of load-bearing structures, the tests provided only indirect information about actual stresses in materials. This was of limited use to engineers wanting to achieve minimum sizes and weights of geometrically complex components used; for instance, in aircraft and other high-performance machines.

While it can be important to know stresses deep within a component under load, the highest stresses are usually found at or near the surface, which is also where cracks begin when materials fail by brittle fracture. Since the early nineteenth century, surface stresses were calculated by measuring the surface strain, for instance the distance between two gauge marks, and inferring the stress knowing the Young's modulus of the material. However, this technique was of little use for components under dynamic loads or with inaccessible surfaces. To meet these challenges, two main strands of technologies for experimental stress analysis were developed during the nineteenth century.

The first made use of the phenomenon of photoelasticity. Certain transparent materials display an anisotropy, called birefringence, to light. When the incident light is polarized, the birefringence has the effect of rotating the plane of polarization of the light. The degree of anisotropy and, hence, rotation, depends on the stresses in the material or, rather, the differences between the stresses in the three principal directions; that is, (σ_1-σ_2), (σ_1-σ_3), and (σ_2-σ_3). If a two-dimensional test piece is used, one principal stress is zero, and the magnitude and direction of the remaining two principal stresses can be established. The only point at which the photoelastic fringes give a direct indication of stress is at the model boundary where a second principal stress (perpendicular to the surface) is also zero. The varying degree of rotation of the incident polarized light is viewed in a polariscope, which displays the highly colorful interference fringes. For quantitative work, monochromatic light (usually sodium) is used to produce sharper fringes (Figure 2).

The early development of this technique was undertaken by E.G. Coker and L.N.G. Filon at University College, London in the early twentieth century using test models made of glass. Although glass is easily available, it is not highly birefringent, is difficult to cut to shape and is vulnerable to fracture. The technique became more accessible with the development of transparent epoxy resins such as Araldite in the 1940s.

The technique reached its peak utility in the 1960s in the aerospace industry, for example at

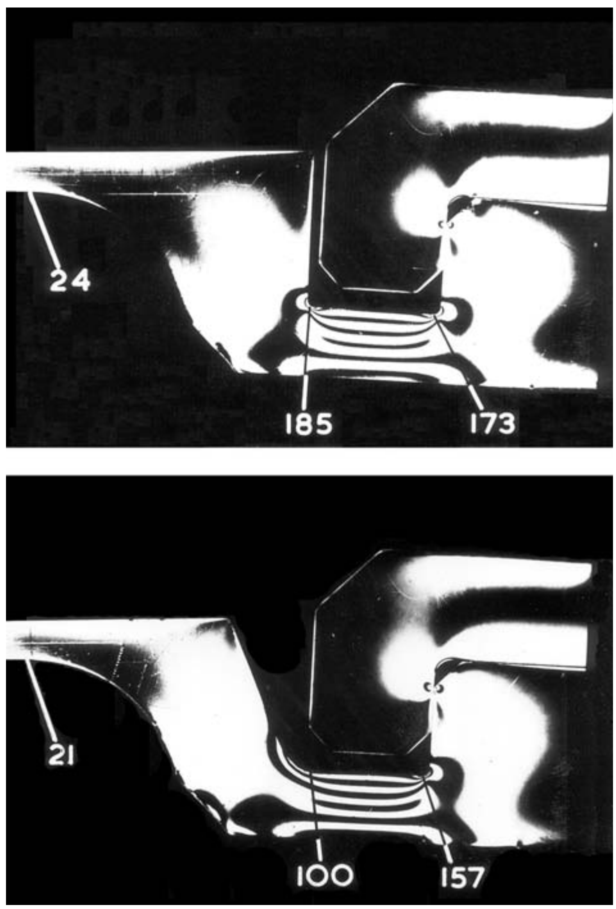


Figure 2. Photoelastic fringes photographed for stress analysis purposes. The comparison of two alternative designs for a clamp in an airengine illustrate a substantial reduction in the peak stresses in Design B.
[*Photos: Rolls Royce Photoelastic Laboratory*]

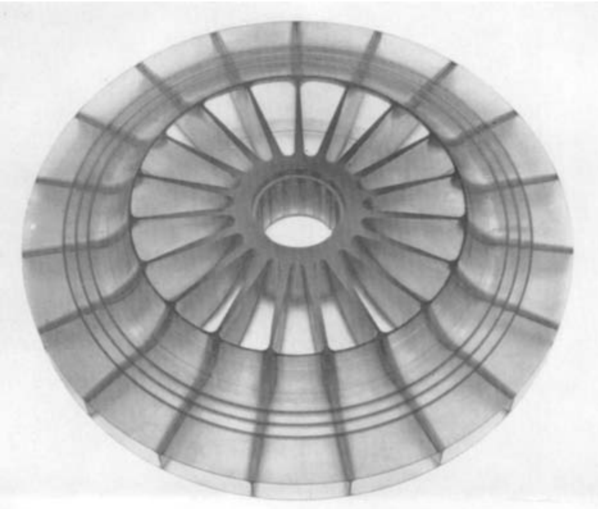


Figure 3. Impellor machined from a solid block of Araldite prior to loading for stress analysis. [*Photo: CIBA*]

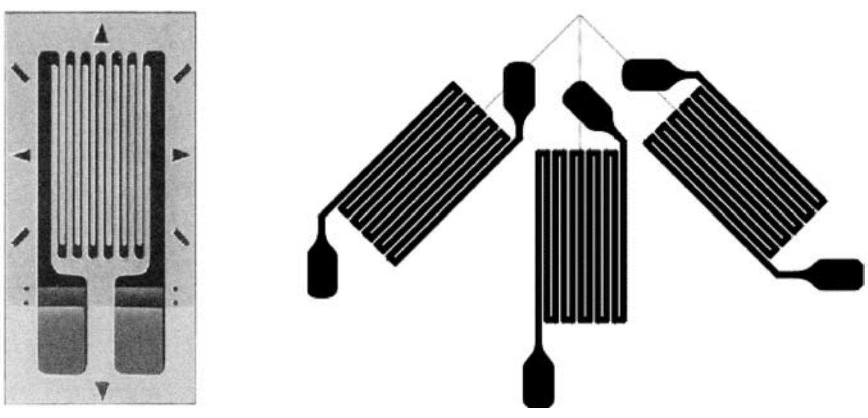
Rolls Royce in Derby, England, where three-dimensional models of impellor or turbine disks were analyzed (Figure 3). A full-size disk, perhaps 600 millimeters in diameter, was intricately machined out of a solid block of Araldite. Dummy turbine blades were fitted. These had a scaled-down weight to take account of the difference in stiffness between the model material and actual material. The entire assembly was rotated at a scaled-down speed, in an oven of about 300°C to allow the loads to cause exaggerated deflections. The oven was then cooled to room temperature with the model still spinning so that the stresses would be frozen or locked into the Araldite. Finally, a number of slices of Araldite, just 1 or 2 millimeters thick, were cut from the model and analyzed using a conventional polariscope.

The second means for measuring surface stresses was the strain gauge, several forms of which were developed. The most widely used depended on the fact that the cross-section of an electrical conductor will reduce when stretched, and this in turn increases its electrical resistance. The change in resistance is usually measured with a Wheatstone bridge. Electrical resistance strain gauges were first used by Charles Kearns in the early 1930s for studying the stresses in aircraft propellers. He ground flat a conventional carbon composite electrical resistor and mounted it on an insulating strip, which he cemented to the blade. He made the electrical connection to the static Wheatstone bridge by means of brushes and rings similar to those used in motors.

In the late 1930s, both Arthur Ruge and Edward Simmons in the U.S. used arrays of fine wires to achieve the same effect. In 1952 Peter Jackson, working with the Saunders–Roe Company on the Isle of Wight, used the new printed circuit board technology to make foil strain gauges that were much smaller, easier to fit, and more reliable. These are still in widespread use, both singly and in the form of a rosette with three gauges at 120 degrees to enable the principal stress directions to be established (Figures 4 and 5).

Despite their simplicity, electric resistance strain gauges had many practical problems, especially the low signal-to-noise ratio of the output. Mechanical brushes were replaced by mercury slip rings, direct current (DC) systems gave way to alternating current (AC) systems, and radio telemetry finally avoided the need for electrical connections. More recently, semiconductor strain gauges provide up to 50 times greater sensitivity. Two remaining problems with all strain gauges is their instability over time and their sensitivity to temperature changes. For the most careful measurement, an electrical thermometer is essential as well as constant recalibration to compensate for creep in the mounting of the gauge on the component under test.

Figure 4. A foil strain gauge and a rosette formation of three such strain gauges. [*Photo: University of Cambridge*]



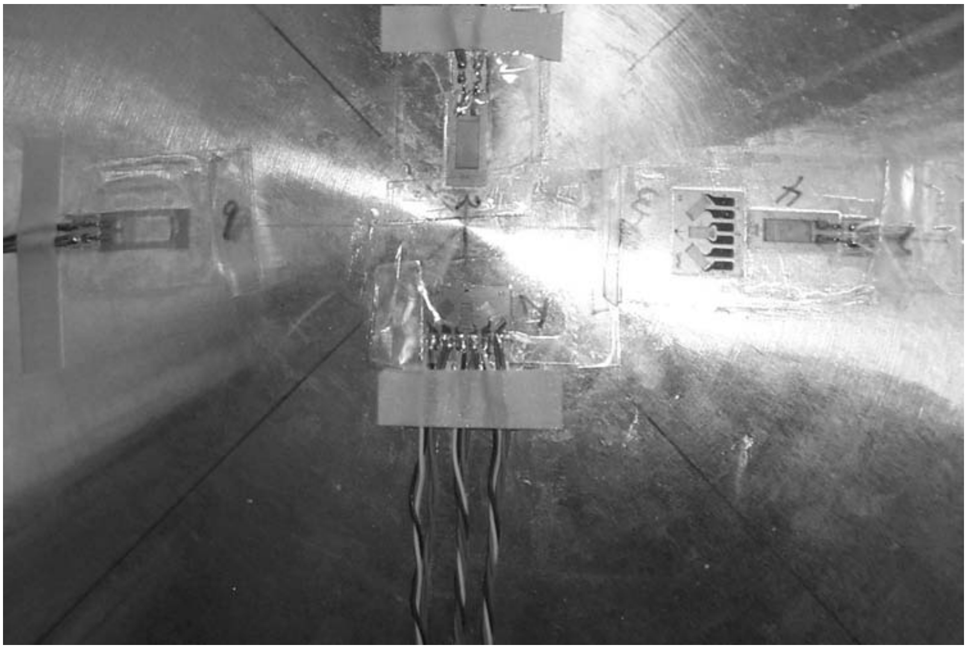


Figure 5. Single strain gauges and rosettes fixed to the surface of the component under test, and wired to measuring device.
[*Photo: Fermilab*]

Another strain gauge technology developed in the 1950s was based on the long-known relationship between the pitch at which a string vibrates and the tension in the string. Vibrating-wire strain gauges were developed mainly for use in large objects such as buildings and bridges, which need to be monitored over long periods of time. A wire, perhaps 200 to 400 millimeters long is fixed at both ends, and its pitch is measured electronically as the length, and hence tension, changes due to relative movement of the two ends. Such gauges are also used in geotechnical investigation of rock movements in mines and bridge foundations.

Two other techniques for studying surface strains were developed in the 1960s. A crude, though very simple technique, was to coat a surface with a brittle lacquer. When strained, the material would crack in a direction perpendicular to the line of maximum principal stress, and the degree of cracking would indicate its magnitude. An advantage of this passive technique was that it could freeze, so to speak, the loading in a moving engine component after the engine had come to rest.

A more sophisticated technique utilized surface photoelasticity. Improvements in epoxy resins in the 1960s allowed them to be fixed directly to the surfaces of components, and the strains induced could be made visible using a reflection polariscope. This technique allowed rapid measurement of strains on complex surfaces and, with the development of holography, also allowed detailed assessment of strains and movement associated with vibration in machinery.

The arrival of optical fibers and lasers during the 1970s led to the development of optical strain gauges. Their advantage over wire strain gauges is the ability to measure the precise point along the length of an optical fiber at which it is stretched. This technology can be used to detect strains in the decks of long bridges.

While photoelasticity and strain gauge technologies have been widely replaced by finite-element stress analysis, they were of vital importance in helping to validate and calibrate such computer models in their early days. Strain gauge and photoelastic coating technologies are both still used when the high costs of setting up computer models are prohibitive.

See also **Bridges, Long Span and Suspension; Bridges, Steel; Buildings, Prefabricated; Concrete; Concrete shells**

W. A. ADDIS

Further Reading

Coker, E.G. and Filon, L.N.G., *Treatise on Photoelasticity*, 2nd revised edn. Cambridge University Press, Cambridge, 1957.

Dally, J.W. and Riley, W.F. *Experimental Stress Analysis,* 3rd edn. McGraw–Hill Education, New York, 1991.

Frocht, M.M. *Photoelasticity*. Chapman & Hall, London, 1941 (vol. 1) and 1948 (vol. 2).
Hendry, A.W. *Elements of Experimental Stress Analysis*. Pergamon Press, Oxford, 1977.
Hetenyi, M. The fundamentals of three dimensional photoelasticity. *J. Appl. Mech*., 5, 4, 149–155, 1938.
Hetenyi, M. Handbook of Experimental Stress Analysis. Wiley, New York, 1950.
Jessop, H.T. and Harris, F.C. *Photoelasticity: Principles and Methods*. Dover, New York, 1950.

Explosives, Commercial

All chemical explosives obtain their energy from the almost instantaneous transformation from an inherently unstable chemical compound into more stable molecules. The breakthrough from the 2000-year old "black powder" to the high explosive of today was achieved with the discovery of the molecular explosive nitroglycerine, produced by nitrating glycerin with a mixture of strong nitric and sulfuric acids. Nitroglycerin, because of its extreme sensitivity and instability, remained a laboratory curiosity until Alfred Nobel solved the problem of how to safely and reliably initiate it with the discovery of the detonator in 1863, a discovery that has been hailed as key to both the principle and practice of explosives. Apart from the detonator, Nobel's major contribution was the invention of dynamite in 1865. This invention tamed nitroglycerine by simply mixing it with an absorbent material called kieselguhr (diatomous earth) as 75 percent nitroglycerin and 25 percent kieselguhr. These two inventions were the basis for the twentieth century explosives industry.

In the years from 1873 to 1920 there was unprecedented development of commercial explosives. The original dynamite was replaced with a range of nitroglycerin-based explosives prepared by mixing varying proportions of nitroglycerin with ammonium nitrate (today well known as a fertilizer) and other nitrates such as sodium and potassium nitrates. Also during this period explosives with the almost household names of blasting gelatin, gelignite, opencast gelignite, and dynamite were developed each designed specifically to achieve maximum efficiency in mining, quarrying, and civil engineering operations.

The introduction of a specialist range of explosives specifically designed to be safe (permitted) when used in dangerous conditions in gassy coal mines, where there is the risk of methane air (firedamp) explosions, occurred from 1922 to 1932. By 1970 compositions had achieved a very high degree of safety even in the most broken strata with the highest danger.

Early detonators were initiated with a length of black powder safety fuse. By 1930 a full range of electric instantaneous and delay detonators had been developed to enable sequential firing, a major advance in blasting efficiency. Today, high-precision blasting is made possible by a range of sophisticated electronic delay detonators.

There are over 200 explosive chemicals in current use, and TNT (trinitrotoluene) is perhaps the best known. TNT was first produced in 1863 by "nitrating" toluene using a mixture of strong nitric and sulfuric acids. The principal use of TNT is in military explosives, and it has been used in shells and bombs since 1902. Shortly after World War I, TNT was used in quarrying and opencast explosives in admixture with potassium nitrate and aluminum. Many other important specialist explosives include: PETN (pentaerithyritol tetranitrate), a crystalline explosive used in detonating transmission lines (Cordtex and Primacord); and RDX (cyclonite trimethylen-trinitramine) used in military applications and the manufacture of boosters for high performance and shaped charges capable of penetrating 3 centimeters of steel plate. These high-performance explosives have helped develop other technologies such as:

1. Metal forming, in which exotic metals such as titanium are precisely shaped by explosively produced hydraulic shock
2. Welding of metal pipe work by shock waves into seamless welds
3. Producing shaped charges that can cut through metal structures such as girders and caissons in demolition, engineering, and oil well perforation

The major twentieth century revolution in commercial explosives technology, however, came in 1956 with the introduction of ammonium nitrate water slurry explosive, a simple mixture of ammonium nitrate and fuel oil (ANFO) containing no explosives, which could be safely mixed on site. ANFO came into prominence following the development in the 1940s of porous ammonium nitrate prills (pellets), which absorbed the oil. The Iron Ore Company of Canada made the first serious use of ANFO in 1956. In the early 1960s the first mobile on-site bulk loading ANFO truck for opencast and quarry blasting was introduced.

The disadvantage of ANFO compositions, however, is that they are not waterproof. This disadvantage was overcome by the development of slurries in the late 1960s. Slurry is a saturated aqueous solution of ammonium nitrate in which porous ammonium nitrate prills and a sensitizing

fuel are incorporated and made water resistant by the addition of thickening agents. In the early days the sensitizer was TNT, but it was soon superseded by nonexplosive aluminum. These new generation slurries were almost immediately put to use in sophisticated on-site mixing trucks that blended and loaded the slurries directly into shot holes. All the components were nonexplosive, and an explosive was only formed when mixed and delivered into shot holes. In 1990 these mobile truck operations were brought into the computer age by incorporating microprocessor control systems to provide programmable loading of shot holes.

In 1968 further development took place with the introduction of emulsion explosives by the Atlas Powder Company in the U.S. Emulsion explosives go one step further in that they do not require any chemical sensitizers whatsoever. These explosives are essentially entrapped air within a water-in-oil emulsion of ammonium nitrate, suspended as microscopically fine droplets and surrounded by a continuous fuel phase of oil and waxes stabilized by an emulsifying agent. The entrapped air acts as a sensitizer and may be in the form of ultrafine air bubbles or air entrapped in glass microballoons. The development of high-efficiency gelled cartridge emulsions in the 1970s made enormous changes, and they are now successfully used in a wide range of underground mining and civil engineering applications. The result of these major developments has been that nitroglycerin-based explosives have been replaced by ANFO and the emulsion explosives in all but the most demanding of hard rock and precision-blasting operations.

Explosives are ideally suited to provide high energy in airless conditions. For that reason explosives have played and will continue to play a vital role in the exploration of space.

See also **Warfare, High-Explosive Shells and Bombs**

JOHN DOLAN

Further Reading

Bergengren, E. *Alfred Nobel: The Man and his Work*. Nelson, London, 1960.

Brown, G.I. *The Big Bang: A History of Explosives*. Sutton Publishing, Stroud, Gloucestershire, 1998.

Cook, M.A. *The Science of High Explosives*. Krieger Publishing, London, 1958.

Croft, W.D. *Dangerous Energy: The Archaeology of Gunpowder*. English Heritage, Swindon, 2000.

Dolan, J. and Langer S.S., Eds. *Explosives in the Service of Man: The Nobel Heritage*. Royal Society of Chemistry, Cambridge, 1997.

Fordham, S. *High Explosives and Propellants*, 2nd edn. Pergamon Press, London, 1980.

Hogg, O.F.G. *Clubs to Cannon: Warfare and Weapons before the Introduction of Gunpowder*. Duckworth, London, 1968.

Meyer, R. *Explosives*, 3rd edn. VCH, Weinhein, 1987.

Urbanski, T. *Chemistry and Technology of Explosives*, 4 vols. Pergamon Press, Oxford, 1984.

F

Farming, Agricultural Methods

Agriculture experienced a transformation in the twentieth century that was vital in increasing food and fiber production for a rising global population. This expansion of production was due to mechanization, the application of science and technology, and the expansion of irrigation. Yet these changes also resulted in the decimation of traditional agricultural systems and an increased reliance on capital, chemicals, water, exploitative labor conditions, and the tides of global marketing.

A sign of the transformation of agriculture in the twentieth century was the shift from China and India as countries often devastated by famine to societies that became exporters of food toward the end of the century. As the world's technological leader, the U.S. was at the vanguard of agricultural change, and Americans in the twentieth century experienced the cheapest food in the history of modern civilization, as witnessed by the epidemic of obesity that emerged in the 1990s. Unfortunately, this abundance sometimes led to overproduction, surplus, and economic crisis on the American farm, which one historian has labeled "the dread of plenty."

Mechanization of agriculture became essential in increasing production of foodstuffs and fiber in the twentieth century. Tractors, powered by internal combustion engines were introduced in the U.S. in the early 1900s. The tractor eventually led to the near extinction of horse- and mule-powered agriculture throughout most of the developed world by the 1950s. The power take-off (PTO) attachment, hydraulics, and three-point hitches were developments in labor-saving equipment that helped prepare the soil and plant and harvest crops. Other mechanical technologies that helped expand acreage and reduce the need for expensive animal and human labor included combine harvesters, refrigerated transportation and cold storage, chemical applicators, cotton, vegetable, and fruit-picking machines, new irrigation technology, as well as machines to test the soil, check the moisture of the crop, and move the crops from field to market.

Science also dramatically transformed agriculture in the twentieth century with the introduction of hybrid crops and genetic engineering of crops. Employing research from the system of land grant colleges such as his alma mater, Iowa State University, Henry A. Wallace (U.S. Vice-President from 1941 to 1945), Roswell Garst, and others commercialized hybrid corn in the 1920s and 1930s. In the 1940s and 1950s hybrid crops combined with petroleum-based fertilizers, insecticides, herbicides, and fungicides led to a "green revolution" in agriculture. From 1970 to 1986, new hybrid rice and wheat strains, fed by artificial fertilizers, led to threefold production in grain yields worldwide.

Western agricultural scientists were often successful in spreading green revolution technology throughout the globe, with mixed results for production, the environment, and indigenous agricultural societies. In the U.S. and around the world, researchers fought for governmental research and development money, and conducted research that could then be applied by farmers able to afford the new technologies. Often the industrialized nations, through the World Bank, the United Nations, or private foundation projects, introduced new technologies in areas where they were not suited or where they severely disrupted

traditional societies or environments. For example, hybrid wheat was introduced into Mexico on a large scale with grants from the Rockefeller Foundation. This wheat was primarily grown to feed livestock, and the hybrid corn introduced in Mexico ended up as fodder because it was not good for making tortillas.

In the field of livestock husbandry for meat production, economies of scale also emerged in the industrialized nations. With fewer farmers, livestock production operations increased in scope. Many cattle, dairy cattle, chickens, and hogs were kept in giant containment facilities that housed thousands of animals that were fed grain, antibiotics, and synthetic growth hormones. Smaller farmers often provided the grain or feeder animals for these large feedlots, dairies, and poultry plants, but more often than not large corporations owned these operations as well. By increasing the size of livestock confinement facilities in many areas, producers had to cope with prodigious amount of manure that they captured in giant sewage lagoons, which often contaminated groundwater supplies and befouled the air for surrounding communities. The meat industry also had to deal with public backlash over the potential health issues of hormone-enhanced meat.

Even with these new technologies, many farmers continued to practice traditional, subsistence type agriculture or to use ancient methods such as flood irrigation in the rice regions of Southeast Asia, planting on terraces in China or Latin America, or slash and burn (swidden) agriculture, in the world's rainforests. In developed countries such as the U.S., new technologies changed agricultural systems, which resulted in the development of single-crop economies, or monocultures, dependent upon large amounts of capital and available land and major inputs of science and technology. In the Great Plains region of the U.S., a traditionally arid region, agriculturists in the 1910s and 1920s practiced "dryland" agriculture. Farmers would intensively plow their new crops, especially right before and after rain. This fell in line with the old adage that "rain would follow the plow." During the years of World War I and for a decade thereafter, wet years resulted in the expansion of planted acreage on the Great Plains, including the phenomenon of "suitcase farmers" who simply planted the crop and returned at harvest time. When the inevitable cycle of dry years returned in the 1930s, the expanded wheat acreage fell victim to drought and the legendary Dust Bowl. After World War II, farmers in the Great Plains region extensively tapped the giant Ogallala aquifer to grow corn and alfalfa in the heart of the former Dust Bowl. In areas where erosion had damaged the soil, farmers also began to practice contour plowing and terrace building as ways of preserving the soil.

Farmers throughout the world were clearly witnessing an end of traditional rural life and agriculture and the rise of global agribusiness. With the introduction of expensive machines, chemicals, and irrigation technology, less efficient, marginal, or smaller-scale farmers and sharecroppers were often forced out of business and off of the land. Governmental policies in many countries supported the removal of smaller farmers to pave the way for larger scale operations. Large multinational corporations and global marketing and transportation industries often stimulated the replacement of traditional agriculture, or made farmers more subordinate as the supplier to large industrial conglomerates. Some agricultural nations in places like Africa and Latin America complained of the unfairness of tariffs and crop subsidies that supported agriculture in developed nations, while wealthier agricultural powers worried about competition from developing countries with their lower labor costs and fewer environmental standards.

In response to the increasing economic, ecological, and human costs of large-scale agriculture, the alternative farming movement emerged in the 1950s and 1960s as a challenge to traditional agricultural systems. While noting that machinery and chemicals had created production increases, organic farmers and their supporters advocated a new "nature-based" model for agriculture rooted in the spirit of E.F. Schumacher's "small is beautiful" concept of technology.

Organic farming and sustainable agriculture systems have several central ideas, including the elimination of chemicals and the use of "natural" fertilizers, such as manure or green manure, or integrated pest management techniques that target bugs with predatory insect releases instead of poisonous pesticides. Organic sustainable farmers also believe in reducing tillage, the use of composts, and even in the mixing of plants (polyculture) to create beneficial ecological interactions between crops, or crops and the soil, pests, or fungi.

Organic farmers also tended to espouse a philosophy that smaller farms and farm communities were of great benefit to society and would be preserved and enhanced by an organic regime. By claiming to be free of harmful poisons and to be more "earth friendly," proponents of alternative agriculture received increased support from con-

sumers and more interest from agricultural researchers who had formerly championed green revolution techniques.

At the end of the twentieth century, trends in agriculture included the rise of the precision farming movement and the growth of aquaculture on a global scale. Precision farming involves age-old farming techniques, in which the farmer is aware of which plots produced well or which tree bore the most fruit and combines this knowledge with a modern systems approach of nearly global knowledge of plant, soil, and weather conditions. Precision farming uses yield monitors (to measure moisture, weight, and other factors), field scouts, sensors, GPS (global positioning systems), and DGPS (differential global positioning systems), and statistical analysis. By having a precise knowledge of growing conditions, precision agriculture allows for the precise application of herbicides, fertilizer, and water, allowing for higher yield, lower long-term costs, and lessened environmental damage. As technology costs declined and more research was invested in the idea of precision agriculture, the movement had an increasing number of adherents, primarily in the U.S.

The revival of aquaculture was another significant trend at the end of the twentieth century. Raising fish or shellfish in confined areas has been practiced for over 4000 years. The Chinese, in particular, have excelled in aquaculture production. Aquaculture can be practiced in saltwater, brackish water, or in freshwater ponds and containers. A minimal amount of grain is required in fish farming as compared to raising beef or pork, and the most popular aquaculture species include Asian carp, tilapia, mussels, catfish, and rainbow trout. At the beginning of the twenty-first century aquaculture was the fastest growing segment of the U.S. agricultural economy, representing over one billion dollars per year. Still, the U.S. lagged behind China, India, and Japan in aquaculture production.

Agricultural systems were greatly influenced by biotechnology, especially with the advent of genetic engineering and the arrival of the first genetically modified crops in the 1980s. The practitioners of animal and plant husbandry have altered nature over the millennia to increase production and profit. But since the advent of Mendelian genetics and the successes of hybridization, the promise and the perils of biotechnology have been a central concern of the agricultural research establishment. Karl Ereky, a Hungarian, coined the term biotechnology in 1919. Many ethicists and environmentalists feared unforeseen consequences of literally "playing God" by splicing genes from one plant's DNA into another plant's DNA, thereby creating a new species. Scientists and farmers noted the great potential to create new strains that increased yield, were pest or drought resistant, or would grow in a particular shape or color. Still, at the end of the twentieth century, a formidable movement, particularly strong in Europe, held steadfast in opposition to genetically modified crops.

With the transportation revolution of the twentieth century, farmers became an increasing component of a global agricultural economy. Globalization of the world economy led to a more rapid adoption of technology and the elimination of traditional agriculture in the developing world. Although many nations such as Japan, the countries of Western Europe, and the U.S. continued to heavily subsidize their agricultural system, global competition increased as tariffs began to fall in the 1980s and 1990s. For example, by the end of the century, even with a 400 percent tariff on Chinese garlic, California growers stopped planting garlic and instead marketed Chinese garlic. A similar process occurred with the apple industry in Washington as lower Chinese labor costs and capital outlays wreaked havoc on competitors. It seems likely that success in feeding the world will continue to require an emphasis on sustainable agriculture and biotechnology.

See also **Breeding, Plant, Genetic; Biotechnology; Farming, Mechanization; Fertilizers; Genetic Engineering, Applications; Irrigation Systems; Pesticides**

Randal Beeman

Further Reading

Beeman, R. and Pritchard, J. *A Green and Permanent Land: Ecology and Agriculture in the 20th Century*. University Press of Kansas, Lawrence, 2001.

Bud, R. Janus-faced biotechnology: a historical perspective. *Trends Biotech*., 7, 1989.

Emmot, A., Hall, J. and Matthews R. The potential for precision farming applied to plantation agriculture. *Proceedings of the 1st European Conference on Precision Agriculture*. Warwick, 1997.

Hurt, R.D. *American Agriculture: A Brief History*. Iowa State University Press, Ames, 1994.

Kloppenberg, J.R. *First the Seed: The Political Economy of Plant Biotechnology*. Cambridge University Press, New York, 1988.

Marcus, A.I. *Cancer from Beef: DES, Federal Food Regulation, and Consumer Confidence*. Johns Hopkins University Press, Baltimore, 1994.

Vasey, D.E. An Ecological History of Agriculture 10,000 B.C.–A.D.10,000. Iowa State University Press, Ames, 1992.

Farming, Growth Promotion

Early in the twentieth century, most farmers fed livestock simple mixtures of grains, perhaps supplemented with various plant or animal byproducts and salt. A smaller group of scientific agriculturalists fed relatively balanced rations that included proteins, carbohydrates, minerals, and fats. Questions remained, however, concerning the ideal ratio of these components, the digestibility of various feeds, the relationship between protein and energy, and more.

The discoveries of various vitamins in the early twentieth century offered clear evidence that proteins, carbohydrates, and fats did not supply all the needs of a growing animal. Additional research demonstrated that trace minerals like iron, copper, calcium, zinc, and manganese are essential tools that build hemoglobin, limit disease, and speed animal growth. Industrially produced nonprotein nitrogenous compounds, especially urea, have also become important feed additives. The rapid expansion of soybean production, especially after 1930, brought additional sources of proteins and amino acids within the reach of many farmers. Meanwhile, wartime and postwar food demands, as well as a substantial interest in the finding industrial uses for farm byproducts, led to the use of wide variety of supplements—oyster shells, molasses, fish parts, alfalfa, cod liver oil, ground phosphates, and more.

By mid-century, researchers had concluded that an unknown "animal protein factor" (APF) was a requirement for maximum animal nutrition and growth promotion. It became increasingly apparent that proteins from animal byproducts better aided livestock growth as opposed to vegetable proteins. University, industry, and government scientists concluded that a nutrient found in liver, labeled Vitamin B12, was the mysterious APF factor. Searching for a method to produce the new vitamin on an industrial scale, manufacturers found a ready solution in residues from the fermentation processes used in the production of the antibiotic "wonder drugs." Meanwhile, researchers began to notice connections between common antibiotics and animal growth. In 1949, E. L. R. Stokstad and Thomas H. Jukes of the American Cyanamid Company found that the Vitamin B12 produced from the antibiotic organism yielded more rapid growth in chicks and hogs than a pure vitamin extracted from liver. Tests using the antibiotic alone, even in subtherapeutic dosages, proved to accelerate animal growth; most initial reports indicated that both chicks and hogs responded to antibiotic feeds, while the impact on ruminants was less. More recently, antibiotics known as ionophores have become commonly used as finishing rations for ruminants. These drugs improve the fermentation efficiency of the rumen, and also reduce bloat, acidosis and other feedlot diseases that are common among animals fattened on grains rather than their natural food of forage grasses.

Hormone supplements also revolutionized animal feeding. In 1954, Wise Burroughs of Iowa State College announced the synthetic estrogen hormone diethylstilbestrol (DES) caused improved feed efficiency on sheep and cattle at very low cost and without reducing carcass quality. Other hormones also became part of animal growth strategies, such as progestins that regulate the estrus cycle and to enhance feed efficiency, and androgens that aid muscle building on the feedlot.

Meanwhile, French researchers F. Stricker and F. Greuter opened up new avenues for research with their 1928 discovery of connections between animal hormones and the milk production of dairy animals. Over the next several decades, scientists gained increasing understanding of the effect of hormones like somatotropin, but isolating natural hormones from pituitary glands of slaughtered animals remained financially infeasible. In the mid-1980s, scientists learned to isolate the cow gene responsible for the growth hormone, remove it from animals, splice it onto bacteria, use bacterial growth to synthesize additional hormones, and inject a purified culture into cows in order to increase both milk production and feed efficiency. More than an incremental step in this history, the use of recombinant DNA techniques represented a revolutionary new development.

Feed supplements are partly responsible for dramatic increases in global meat production and consumption, especially since World War II. Antibiotics and growth hormones can help speed animal growth, improve feed efficiency, increase the number and viability of eggs and animal offspring, and reduce animal diseases. In the competitive marketplace of western agriculture, most farmers quickly embraced them. Manufactured and medicated feeds spurred farmers to increase the size and capital investment of livestock operations, to manipulate the natural rhythms of animals' breeding, birth, weaning, rebreeding, and slaughter, and to move livestock

into ever more confined, streamlined, and centralized operations.

These growth-promoting agents have generated controversy. Concerns about antibiotic resistance have been voiced since the early 1950s, and various studies have indicated plausible links between animal antibiotics and human disease. Although repeated efforts to regulate the use of antibiotics in subtherapuetic dosages in the U.S. have stalled repeatedly, other nations have restricted their usage. In 1954, officials in the Netherlands announced a ban on all antibiotic-enhanced feeds. In the U.K. in 1969, a special commission recommended that any antibiotics used for growth should be different from those used in human or animal medical care. Pharmaceutical firms quickly replaced these drugs with alternative antibiotics, yet several of these in turn have been banned throughout the European Union, effective in 1999.

The debate over growth hormones has been even more contentious, and agricultural feed additives have been at the center point of social, political, and ethical debates over bioengineering. News of the "DES daughters" broke in 1970. These seven young women had developed rare a form of vaginal cancer that could be traced to their mothers' use of DES as a drug intended to prevent miscarriages. The U.S. finally banned the drug in 1979, after more than two dozen other nations had already done so. Although other livestock hormones remained legal (or otherwise available) in the U.S., European policymakers and consumers both have been increasingly stringent in enforcing regulations. In recent years, even the name of the hormone technology has become controversial; opponents prefer the term recombinant bovine growth hormones, (rBGH), while supporters prefer to use the term recombinant bovine somatotropins (rbST). The BGH/rbST debate has moved beyond questions about medical or environmental objections to genetic engineering. For many, the issue is the social cost of allowing enhanced milk production technologies to eliminate many family dairy farms.

In response to some of these controversies, niche markets have emerged for livestock that are free from confined environments and for animals that are antibiotic and growth hormone free. Although such trends were evident at the end of the twentieth century, particularly in western and central Europe, it remains unlikely that such protests will slow the penetration of manufactured and medicated animals feeds into the developing world.

See also **Antibiotics, Use after 1945**

MARK FINLAY

Further Reading

Bud, R. *The Uses of Life: A History of Biotechnology*. Cambridge University Press, Cambridge, 1993.

Feed Additives: The Added Value to Feed. Nefato, Aalsmeer, Netherlands, 1998.

Johnson, A. *Factory Farming*. Blackwell, Oxford, 1991.

Jukes, T.H. *Antibiotics in Nutrition*. Antibiotics Monographs, 4, Medical Encyclopedia Inc., New York, 1955.

Krimsky, S. and Wrubel, R. *Agricultural Biotechnology and the Environment: Science, Policy, and Social Issues*. University of Illinois Press, Chicago, 1996.

McCollum, E. *A History of Nutrition: The Sequence of Ideas in Nutrition Investigations*. Houghton Mifflin, Boston, 1957.

Marcus, A.I. *Cancer from Beef: DES, Federal Food Regulation, and Consumer Confidence*. Johns Hopkins University Press, Baltimore, 1993.

Sammons, T.G. Animal feed additives, 1940–1966. *Agricult. Hist.*, 42, 305–313, 1968.

Farming, Mechanization

Mechanization of agriculture in the twentieth century helped to dramatically increase global production of food and fiber to feed and clothe a burgeoning world population. Among the significant developments in agricultural mechanization in the twentieth century were the introduction of the tractor, various mechanical harvesters and pickers, and labor-saving technologies associated with internal combustion engines, electric motors, and hydraulics. While mechanization increased output and relieved some of the drudgery and hard work of rural life, it also created unintended consequences for rural societies and the natural environment.

By decreasing the need for labor, mechanization helped accelerate the population migration from rural to urban areas. For example, in 1790, 90 percent of Americans worked in agriculture, yet by 2000 only about 3 percent of the American workforce was rural. Blessed with great expanses of land and limited labor, technologically inclined Americans dominated the mechanization of agriculture during the twentieth century. Due to mechanization, irrigation, and science, the average American farmer in 1940 fed an estimated ten people, and by 2000 the number was over 100 people. Yet even as mechanization increased the speed of planting and harvesting, reduced labor costs, and increased profits, mechanization also created widespread technological unemployment in the countryside and resulted in huge losses in the rural population.

After Nicholas Otto patented the internal combustion engine in 1876, the days of horse and

steam power in agriculture were eventually eclipsed on the farm by gasoline and diesel engine power. Early steam tractors weighed as much as 5 tons or more and required substantial labor to operate and fuel. The operators were continually at risk from possible explosion of the boiler that powered these iron-wheeled beasts. A number of American manufacturers, such as the Hart-Parr Company, the John Deere Plow Company, and International Harvester, were engaged in building gasoline-powered tractors at the beginning of the twentieth century.

As tractors became more powerful, practical, and affordable, their presence increased on the farmstead. With a tricycle design for enhanced maneuverability, a PTO (power take-off) to run equipment, hydraulic lifts, a three point hitch, and pneumatic tires, tractors could pull a plow, mow and bale hay, disk weeds, or dig furrows among other tasks. In 1907 some 600 gasoline-fueled tractors were in operation in the U.S.; by 1950 nearly 3.5 million tractors were in use in the U.S. alone.

Tractors, unlike horses and mules, did not require five acres to feed every year, and they did not need rest. Henry Ford's all-purpose Fordson in 1917, and the International Harvester Farmall in 1924 enjoyed high sales in the 1920s. Other tractor improvements included roll bars, improved seating, variable hydrostatic transmission for increased speeds, and Harry Ferguson's hydromechanical servo, which allowed the tractor operator to control the depth of implements in the soil by keeping a constant load on the tractor. The first diesel designed to power movable machinery in the U.S. was built in 1930, and by the mid-1930s the first diesel tractor was sold.

From the World War I campaign to increase wheat production in the U.S. to the Soviet Union's "virgin lands" projects of the 1950s and 1960s, tractors have increased the amount of land under cultivation throughout the globe. While yields increased, so did potential environmental catastrophes, such as the American Dust Bowl in the 1930s, and the failure of much of Soviet's virgin lands program. By the end of the twentieth century the U.S. had only two major tractor manufacturers, and India led the world in tractor production, building nearly 300,000 units per year. In Africa most of the farm labor continued to be done by people or animals.

Along with the gasoline-powered tractor came other devices to increase efficiency and profits by reducing labor and expanding the scope of production. In the large wheat area of the American West, the first combine harvesters appeared in the 1890s. After Cyrus McCormick's famous reaper appeared in the 1830s, inventors tried to find ways to combine the harvesting and threshing process. Hiram Moore and others made several attempts at manufacturing combines in the mid-nineteenth century, but it took until around 1900 for large-scale farmers on the American West Coast to employ the first commercially and mechanically viable combines.

Tractors pulled the first combines, and they used a rotary header to gather the plant, a cutting bar, and an internal thresher, but by the 1940s the machines were self-propelled. Gasoline powered engines or the PTO of a tractor operated the cutter, and the various mechanical fingers and pulleys that drew the crop into the thresher and shook the grain loose from the plant, with the grain being augured into a grain tank, and the crop residue sprayed back onto the field. Early combines required a four-man crew and could cut up to a 16-foot (5-meter) swath. Over the years combines became larger, with rubber tires, self-enclosed air conditioned cabs, headlights to allow night harvesting, and wider swathes.

Combines represented a tremendous capital investment for farmers, and they contributed to the growth in the size of individual farms and the population decline in rural communities. But whether a farmer purchased a combine, or hired a custom crew, combines were less expensive and more dependable than hiring a crew of laborers to harvest large fields. While the first combines primarily harvested wheat, mechanical engineers would later build machines to combine harvest rice, oats, soybeans, and corn by mid-century. Early models of these machines often left up to 50 percent of the grain in the field and were prone to problems when operated in muddy fields as well as mechanical difficulties. Yet even with these difficulties, combine harvesters became increasingly efficient, and they remain a symbol of the mechanization of agriculture in the twentieth century.

Inventors, corporations, mechanical engineers, and professors perfected machines to harvest grains, but cotton remained primarily a hand-picked crop until the 1950s. Cotton bolls ripened at different times, and developing a machine that could pick trash-free fiber without damaging the cotton took over 100 years. In the 1920s and 1930s, early versions of the cotton picker often used vacuum pumps and blowers to either suction or blow the cotton into large bags. Engineers John and Mack Rust experimented with combines that

Figure 1. Mechanical corn picker, 1939, Grundy County, Iowa.
[*Courtesy U.S. Department of Agriculture.*]

used smooth wet spindles to pick the crop and barbed and serrated spindles to twist the lint from the boll. In the mid-1930s the Rust machine could pick almost 180 kilograms of cotton per hour. In 1948 the International Harvester Corporation was the first company to successfully market a mechanical cotton picker. It used rotating metal drums with wetted spindles that pulled the lint from the bolls, rubber doffers to remove the fibers from the spindles, and an air conveyor to blow the cotton into a container.

Cotton harvesters became more efficient, less expensive, and more economically viable for farmers, especially those with larger holdings. Additional improvements reduced the amount of plant material, or trash, in the cotton, and by the late 1960s most of the cotton crop in the U.S. was machine picked. An early two-row cotton picker displaced 80 workers, and in the U.S. alone millions of workers were technologically unemployed by the mechanization of the cotton harvest.

Like cotton, sugar beet harvesting was difficult to mechanize. Sugar beets have deep roots and heavy foliage. Growers relied on a large number of laborers to grow, harvest, and transport the crop. Beets had to be pulled from the ground, topped with a knife, and then loaded onto trucks or wagons. In the 1930s it took about 30 hours of labor to harvest an acre of sugar beets. A successful harvesting device would require the machine to top the beets and to remove the excess dirt and clods from the crop. Early machines topped the beets at ground level and lifted the roots with a spiked wheel onto a conveyor belt and into a truck. By the 1970s sugar beet harvesters had been enhanced to the point where they could harvest 24 acres in a ten-hour day.

Mechanization of agriculture most often occurred in developed nations willing to invest in labor-saving technology. In the U.S., farmers in the Central Valley of California have been at the forefront of agricultural mechanization. The early combines and cotton pickers were first utilized in California, where growers were often dependent upon migratory labor. Tomatoes were another major California crop that required vast supplies of labor. Creating a mechanical tomato picker involved a number of obstacles. Tomatoes on the plant ripened at different times, the vines snarled machines, and the fruit itself was easily bruised and not suited to machines. Consequently, to develop the machine-harvested tomato required the cooperation of engineers and crop scientists. New varieties of tomatoes that ripened simultaneously and were tough skinned emerged from university and corporate laboratories and test plots, such as

the work done by G. C. Hanna at the University of California, Davis.

The tomato picker developed by the 1960s worked by lifting and cutting the vines and shaking the plant on a shaking bed, where the harvested tomatoes would then fall onto a conveyor belt which transported the fruit to a bin while the vines fell to the ground. Though the early pickers had many technical problems and often malfunctioned, they could harvest 10 tons, or one third of an acre of tomatoes, per hour. Most of the tomatoes harvested mechanically ended up as catsup, juice, or tomato paste, and the process of mechanically harvesting market-fresh tomatoes remained problematic in the 1990s. Tens of thousands of farm laborers in California alone lost their jobs with the advent of tomato harvesters. The mechanical tomato harvester also allowed for a sweeping increase in tomato production worldwide.

When contemplating the mechanization of agriculture, we associate large and complicated machines roaring through a field as the symbol of technological change in the countryside. But mechanization of the farm had many smaller, less extravagant elements, such as the influence of electrification. Many rural communities in the developing world still lack electricity. In the U.S. in the early 1930s, the federal Rural Electrification Administration helped bring electricity to the farms of America. Electricity powered milking machines, cream separators, and irrigation pumps. Electric power also allowed for environmentally controlled barns and greenhouses, cold storage units, and grain dryers that facilitated the mechanization of the grain harvest. Small electric motors could make a number of tasks on the farm easier, whether it be an auger to move grain, or a grinder or arc welder in the workshop.

Many other devices helped to mechanize agriculture in the twentieth century. Automobiles, trucks, and other transportation improvements transformed rural society and created new market opportunities and new competitors for farmers. Giant feed lots confined thousands of animals for meat or dairy production using highly mechanized facilities for feeding the animals and removing the waste.

Citrus growers used wheeled machines with revolving blades to trim the trees in their orchards; in almond orchards, machines grab and shake the fruit from the tree and another machine sweeps the crop from the ground. Pesticide sprayers with tall wheels and giant booms glide over rows of vegetables. On grape vineyards plastic-fingered pickers shake grapes from the vines on their way to be pressed into wine; and elaborate systems for drip irrigation allow for the precise application of water, fertilizers, and other chemicals.

Every year brings new changes, technological improvements and mechanical innovations, but with each new step toward mechanization there is a corresponding increase in capital costs, technological unemployment, and environmental disruption. Clearly the move from human- and horse-powered agriculture reduces the number of people required to grow food and fiber, and by the last half of the twentieth century developing countries were embracing the mechanization of their indigenous agriculture on the Western model.

See also **Dairy Farming; Farming, Agricultural Methods; Irrigation Systems**

RANDAL BEEMAN

Further Reading

Hurt, R.D. *Agricultural Technology in the Twentieth Century*. Sunflower University Press, Manhattan, 1991.

Marcus, A.I. *Technology in America: A Brief History*. Harcourt-Brace, San Diego, 1989.

Rasmussen, W. The mechanization of agriculture. *Scientific American*, September, 1982.

Fax Machine

Fax, or facsimile, technology refers to the concept of replicating printed documents across long distances and dates back to the nineteenth century, along with the advent of the telegraph. Embracing the emerging electromechanical technology of the time, several devices were developed; however diffusion was limited due to the elaborate and intricate mechanism required as well as interoperability between disparate devices. Among the first on record was Alexander Bain's "chemical telegraph," patented in America in 1843. The device used a metallic contact that sensed the raised text, which triggered the flow of electric current. In 1847 Frederick Bakewell invented the "copying telegraph," which introduced the concept of scanning the source document line by line. Both systems required pendulums and electromagnets for synchronization. However, fierce competition with Samuel Morse over long-distance telegraph lines led to legal disputes and Bain's patent was declared invalid. Although the facsimile technology did not proliferate along with the telegraph, the concept of facsimile transmission did, and isolated systems continued to emerge wherever telegraphic networks were set up. A successful system was demonstrated in France in 1862 by Abbé Casselli,

and a network of commercial stations was established. In America, Elisha Gray developed a system comprised of rheostats and electromagnets. Despite some successes, the largely mechanical technology was cumbersome, and lack of international interoperability and slowness of the system compared to Morse's telegraph prevented commercial success and widespread proliferation.

At the beginning of the twentieth century, the science of reducing images to scanned lines of information which could be electrically transmitted for distant reproduction was maturing but the technology was slow to follow. In 1902 Otto von Bronk obtained a patent for the principle that an image could be constructed out a series of lines. Einstein's mathematicalization of the photoelectric phenomena and other developments in quantum physics in the early twentieth century provided the necessary scientific understanding for the emergence of photoelectric technology, which was primarily based on the element selenium. Dr. Arthur Korn invented an improved and practical fax in 1902 based on this photoelectric technology. It should be noted the development of the cathode ray tube television shared the same technological roots as the fax machine; that is, image transfer by sequentially scanned lines.

Early attempts to distribute printed images to domestic consumers in America did not find much success. However, the need to rapidly transmit photographic images for publication in newspapers and elsewhere drove the technology of the fax machine into elaborate commercial products. Along with telegraphic and telephone networks, long-distance faxing started to make use of short-wave radio communications as well. By 1910 Korn had established a network with images stations in Berlin, Paris and London, and in 1922 successfully transmitted a picture from Rome to New York by radio. In 1926 a commercial radio link for facsimile image transmission was operational between London and New York. The limitation of this system was that original image had to be a negative. Subsequently Edouard Belin developed the "Belinograph" in France in 1925. The image to be transmitted was placed on a cylinder and scanned with a powerful light beam and a photoelectric cell converted the reflected light into transmittable electrical impulses.

The American Telephone & Telegraph Company (AT&T) produced a "telephotography" machine in 1924 which was successfully used to send photos over long distances for newspaper publications. In 1934, the Associated Press news agency introduced a similar system for transmitting "wire photos" between centers. These devices perfected the spinning drum and moving photoelectric cell concept, but required mechanical precision and specialist upkeep. As their predecessors did, a continuously varying electrical signal was used to transmit the information. The limitation was that this was subject to decay and electromagnetic interference and noise along the path of transmission, although various signal modulation techniques were used to increase fidelity and resilience. This technology continued to be used in press rooms until the arrival of personal computers and the means of digitalization of images later in the century.

Other implementations of facsimile technology in this period were for public use; the "phototelegram" service as Western Union called it. Facsimile machines were made available in public places in the 1930s for the transmission of messages although this was short lived. Facsimile technology also proved successful in augmenting international telegraph and later, telex facilities. International public fax services grew from the initial New York–London link in 1926 to 24 countries by 1950 and 65 countries by 1976. Largely these facilities were part of the public services offered by the postal services.

The American company Xerox had successfully developed the commercial photocopier in 1959 and made a global commercial success out of it. The Long Distance Xerograph (LDX), which was announced in 1964, stemmed from photocopier technology but was still a cumbersome, expensive and difficult-to-operate device. In 1966, the smaller Xerox Magnafax Telecopier operated on ordinary telephone lines, and was easier to use but still took 6 minutes to transmit a single page. At the Xerox Corporation research institute, the Xerox Palo Alto Research Center (PARC) some fundamental developments and innovations were taking place in digital technology in an atmosphere that was described as "pure invention." Several factors were driving the digitization of the fax, one being that digital information could be compressed using mathematical algorithms (by up factors of up to ten times) and once digitized, the information was not subject to decay or electrical interference during transmission.

The first digital fax machine was built by Lynn Conway at the PARC in 1972 using discrete logical components. The Sierra was a large power-hungry unit that was more of a proof-of-concept than any sort of marketable device. However digital consumer devices were being made possible by the miniaturization of digital electronics (the inte-

grated circuit led to the large-scale integration of digital logic components into the microprocessor and the RAM memory chip, key components of the microcomputer).

International standardization of facsimile communication protocols played a vital role in the widespread adoption of the fax machine since disparate standards limited diffusion. The International Telegraph and Telephone Consultative Committee (CCITT) agreed on the Group 1 standard in 1968. A page took 6 minutes to transmit and was reduced to a series of dots which were either "on" or "off." This sequence was used to encode a signal using frequency shift keying (FSK) modulation technique. Subsequent signal processing enhancements led to the Group 2 standard in 1976, which halved the transmission time. Embracing the burgeoning digital technology of the late 1970s, the Group 3 CCITT standard was agreed to 1980, utilizing the modified Huffman coding for compression of data, which led to a transmission rate of less than 1 minute per page with an improved resolution. This standard also incorporated well-established computer data modem communication protocols, including the facility to "step-down" or scale the data communication rate according to the quality of the analog telephone line connection. This made the Group 3 fax versatile and fault-tolerant. The Group 4 standard came out in 1984 and was orientated around digital ISDN telephone services, however ISDN did not proliferate as expected and Group 3 became the most widely used standard internationally.

The impetus fell to the large Japanese consumer electronic manufacturers such as NEC Corporation and Matsushita Communications in the early 1980s to mass-produce digital fax machines. Electronic developments had enabled an easy-to-use compact solid-state device to be mass produced, and with large volumes and international compatibility, the fax machine became affordable and globally ubiquitous within a decade. Parallel to this was the development of the personal computer, and by 1985 fax peripherals were available to personal computer users which enabled faxing direct from a software application. This made "broadcast" faxing possible, whereby one message could be sent to many recipients with relative ease. Thus the fax became a mass one-too-many marketing and communication tool.

Fax technology was especially useful for international commercial communication, which was traditionally the realm of the Telex machine, which only relayed Western alpha-numeric content. A fax machine could transmit a page of information regardless of what information it contained, and this led to rapid and widespread adoption in developing Asian countries during the 1980s. With the proliferation of the Internet and electronic e-mail in the last decade of the twentieth century, fax technology became less used for correspondence. At the close of the century, the fax machine was still widely used internationally for the transmission of documents of all forms, with the "hard copy" aspect giving many a sense of permanence that other electronic communication lacked.

See also **Electronic Communications; Photocopiers; Telephony, Long Distance; Television, Electro-Mechanical Systems; Printers**

BRUCE GILLESPIE

Further Reading

Coopersmith, J. The failure of fax: when a vision is not enough. *Econ. Bus. Hist.*, 23, 272–282, 1994.

Hiltzik, M. *Dealers of Lightning. Xerox PARC and the Dawn of the Computer Age*. Orion Business, London, 2000.

Petersen, M.J. The emergence of a mass market for fax machines. *Technol. Soc.* 17, 4, 469–82, 1995.

Winston, B. *Media, Technology and Society, A History: From the Telegraph to the Internet*. Routledge, London, 1998.

Feedstocks

The word feedstock refers to the raw material consumed by the organic chemical industry. Sometimes, feedstock is given a more restricted meaning than raw material and thus applied to naphtha or ethylene, but not petroleum. The inorganic chemical industry also consumes raw materials, but the feedstock tends to be specific to the process in question, such as sulfur in the case of sulfuric acid. The development and growth of new feedstocks has driven the evolution of the organic chemical industry over the last two centuries. To a large extent, the history of this industry is the history of its feedstocks. Until the nineteenth century, the only significant raw material for the nascent organic chemical industry was fermentation-based ethanol (ethyl alcohol). Gradually, the products of wood distillation also became important, only to be overshadowed after 1860 by the coal-tar industry. As the organic chemical industry expanded in both size and scope between 1880 and 1930, the need for new feedstocks became urgent. The competition between coal and petroleum was resolved in favor of the latter in the late 1950s. The

petrochemical industry has been phenomenally successful, underwriting the postwar boom in organic chemicals and plastics and weathering the oil crises of the 1970s with minimal damage. Its long-term sustainability remains an issue, and increasing attention is being paid to renewable feedstocks.

Fermentation-Based Chemicals

In dilute aqueous form (beer, wine), fermentation-based ethanol has been known for thousands of years. Pure ethanol was well known by the thirteenth century, but there was only a tiny demand, mainly for pharmaceuticals. Around the beginning of the twentieth century, it became possible to produce lactic acid by fermentation of potato starch. A few years later, Chaim Weizmann at the University of Manchester working in collaboration with the consulting firm of Strange and Graham developed a fermentation process for the production of butanol and acetone from carbohydrates obtained from grain or potatoes. Initially created to produce butanol for synthetic rubber production, it became more important as a source of acetone during World War I. In the early 1920s, the strong demand by the automobile industry for solvents, favored its co-product, butanol. To meet the American demand, the Commercial Solvents Corporation was established at Terre Haute, Indiana, to operate the Weizmann process using maize as the starting material. By 1924, the plant was producing 6500 tons of butanol. In the late 1920s, Friedrich Bergius—better known for his research on the conversion of coal to oil—developed a method of hydrolyzing wood with acids to produce sugars. The process was used in England for a while, and Bergius set up a plant in Germany, but it was not a commercial success. The German firm I.G. Farben did consider using the process to make synthetic rubber (via fermentation-based ethanol) but stayed with coal-based chemistry. The idea of using ethanol as a feedstock—from wine surpluses rather than wood—was revisited by the West German synthetic rubber producer Chemische Werke Hüls in the early 1950s, but again this route was not pursued.

Wood Distillation

Although it had been known since the seventeenth century that heating wood strongly produced wood spirit and pyroligneous acid, the industry concentrated on the production of wood tar and charcoal up to 1800 or so when pyroligneous acid was identified as acetic acid. By 1820, the industry existed in Britain (mainly in Scotland), Germany and Austria. Wood distillation in America started about 1830. During this period, acetic acid, used to make mordant salts and pigments, was the key product. The commercial production of methanol and acetone only began in the 1850s. The importance of wood distillation should not be underestimated. HIAG (Holzverkohlungs-Industrie AG) was the 77th largest German company in 1913, but it declined in the 1920s and was taken over by Degussa in 1931. This was partly a result of the introduction of new methods of making its core products, acetone and methanol.

Coal-Tar

The coal-tar industry was a byproduct of the expanding coal-gas industry, and the tar that was originally wasted became a valuable commodity. When the dry distillation of coal to produce coal-gas for lighting and heating was developed in the early nineteenth century, the coal-tar waste product was largely seen as a noxious nuisance to be burned in open pans or dumped in nearby rivers. The value of coal-tar naphtha as a solvent was soon recognized, and by the 1840s chemists had also isolated the so-called aromatic chemicals from coal-tar including benzene and phenol. Although benzene was quickly taken up as a powerful solvent, it was the development of the synthetic dye industry following William Henry Perkin's discovery of mauve in 1856 that drove the growth of the coal-tar industry. Within a decade, coal-tar was being converted into numerous chemicals by distilling chemicals from the crude tar and subsequent chemical treatment such as nitration. By the 1880s, coal-tar was also used for pharmaceuticals as well as dyes. It is an indication of the former importance of coal-tar that the first major synthetic fiber—nylon—was made from coal-tar chemicals until the 1950s in the U.S. DuPont introduced a process that used furfural, a chemical obtained from maize cobs, in 1948 and only adopted a petrochemical route based on butadiene in 1961.

Coal

Coal itself was not used as a raw material for organic chemicals until the twentieth century. Its exploitation was a combination of supply and demand. In the late 1890s, entrepreneurs had hoped to develop acetylene obtained from calcium carbide as a new method of lighting. Although acetylene gas lighting was used for many years, the

entrepreneurs' hopes were largely misplaced, and there was soon a surplus of calcium carbide. This surplus was partly consumed by the manufacture of a novel fertilizer calcium cyanamide, but acetylene was now available as a versatile and relatively cheap feedstock for organic chemicals. The German chemical industry, chiefly Hoechst and Wacker, began with the manufacture of chlorinated hydrocarbons and acetic acid, which in turn was converted into acetone. At this point, at the outset of World War I, the Haber–Bosch process for nitrogen fixation energized the production of carbon monoxide from coke. (In the Haber–Bosch process, this carbon monoxide was an intermediate in the production of synthesis gas, a mixture of hydrogen and nitrogen.) In 1923, Badische Anilin and Soda-Fabrik (BASF) started the production of synthetic methanol by reacting a mixture of carbon monoxide and hydrogen under pressure. The two lines of approach—acetylene chemistry and high-pressure chemistry—were brought together by BASF's Walter Reppe, who developed a number of interesting coal-based processes.

Acetone

Acetone was one of the first major organic chemicals. It was produced by wood distillation, but the yield was low. In the mid-nineteenth century it was discovered that acetone could be made by heating calcium acetate (from acetic acid, a more abundant product of wood distillation). This "grey acetate" was exported from the U.S. to Germany. By the end of the century, acetone had become an important component of cordite smokeless powder manufacture (as a solvent for gun cotton). During World War I, the Germans developed the production of acetone from acetylene, via acetaldehyde. This process gave a major impetus to the development of acetylene chemistry. By contrast, the British relied on the Weizmann fermentation process. The petrochemical production of acetone from propylene began at Union Carbide's South Charleston, West Virginia, plant in 1928, but the production of acetone from coal-based acetylene continued in West Germany until 1963.

The Automobile

In the early 1920s there was an unprecedented demand for organic chemicals from the booming automobile industry. Car production used nitrocellulose-based lacquers, which required organic solvents. A key compound was butanol, which was initially made by the Weizmann fermentation process. BASF then developed a method of making butanol from acetylene. In the winter, motorists needed antifreeze. Antifreeze was initially based on methanol or the more expensive glycerol, but ethylene glycol (Prestone) was introduced by Union Carbide in 1927. This was one of the first organic chemical products to be made from natural gas.

Synthetic Rubber

The first synthetic rubbers were made in the first decade of the twentieth century, and industrial production was carried on in Germany during World War I. Large-scale production did not take place, however, until the mid-1930s in Soviet Russia and somewhat later in Nazi Germany. The U.S. developed a massive synthetic rubber industry during World War II. Most synthetic rubbers are based on butadiene, but it can be produced from a variety of feedstocks. The early "methyl rubber" was made from acetone, which in turn could be wood- or coal-based. Nazi Germany (and subsequently East Germany) used a lengthy synthesis based on acetylene obtained from coal via calcium carbide or by cracking natural gas. The huge quantities of acetylene needed to make synthetic rubber effectively created an "acetylene chemical industry" in Germany in the 1940s and 1950s. The Soviet industry used the Lebedev process, which converted ethanol into butadiene in one step. The German industry admired the Lebedev process and on two occasions considered switching to it. With the support of Chaim Weizmann, the midwestern farm lobby in the U.S. also promoted this process, and 82 percent of U.S. butadiene was produced from ethanol in 1943. Of far greater importance, both at the time and in terms of its eventual impact, was the development of petroleum-based routes to butadiene. Not only did petroleum account for 59 percent of the U.S.'s butadiene in 1945, it also accelerated the entry of the petroleum refiners into the chemical industry.

Petrochemicals

The petrochemical industry was a result of the massive increase in petroleum refining in the twentieth century. There were two reasons for this. As petroleum refining became more sophisticated, refiners used chemical processes to produce better gasoline (petrol). At the same time, the refineries were producing large quantities of hydrocarbon gases, which were originally burned off. In

the 1930s, the American industry developed methods of making various chemicals from these gases including propanol (Union Carbide), isobutanol (Shell) and styrene (Dow). Other companies such as Jersey Standard (now Exxon) and Anglo–Iranian (British Petroleum, or BP) moved into petrochemicals in the 1930s, partly because of the contemporary interest in the hydrogenation of oil and coal. Petrochemicals is a convenient term because it combines chemicals from petroleum with chemicals from natural gas, a distinction which is all too rarely made. One of the earliest uses of natural gas in the chemical industry was the production of acetylene from Dutch natural gas in the German synthetic rubber industry by passing it through an electric arc. Subsequently it was used to replace coal in the Haber–Bosch and synthetic methanol processes. Natural gas is more important in the American chemical industry than in Europe.

Polyethylene

Imperial Chemical Industries (ICI) introduced polyethylene in 1938, and by 1962 British production had reached 155,000 tons. Just as synthetic rubber had generated a vast demand for butadiene, the expansion of polyethylene manufacture after World War II created an unprecedented demand for ethylene. When ICI first made polyethylene, it used fermentation ethanol as its raw material, thanks to its links with the Distillers Company. It soon became clear however, that this source of ethylene was both too expensive and too limited to meet the soaring demand. In Germany, the main source of ethylene was coke-oven gas, which was also in relatively short supply. During World War II, the Germans even made ethylene by adding hydrogen to acetylene. After the war, it was clear that the petrochemical route was the only one that could provide ethylene cheaply and in the quantities needed. Just as synthetic rubber had boosted the use of acetylene in the prewar German chemical industry, the production of petroleum-based ethylene for polyethylene acted as the vanguard for the petrochemical industry, especially outside the U.S.

Replacing Coal

The new petrochemical industry did not spring up overnight. The established chemical factories were based on older feedstocks, and their chemists had developed a deep understanding of the various processes they operated. Switching to petrochemicals meant building new plants, developing new processes, learning a different kind of chemistry, and even changing the location of production. Vinyl chloride monomer for the manufacture of polyvinyl chloride (PVC)—hitherto made by adding hydrogen chloride to acetylene—was now produced by heating ethylene dichloride. The older factories housed in brick or concrete buildings were replaced by plants with shiny piping and tall column reactors sited in the open air. It is therefore not surprising that the changeover to petrochemicals took two decades in Western Europe and even longer in the Communist bloc. Petrochemical processes worked better for many industrial chemicals such as ethylene, butadiene, or acetone, but other important chemicals such as terephthalic acid (for polyesters) or aniline (for dyes) were more naturally made from coal. In order for petrochemicals to replace coal altogether, it was necessary to produce the so-called aromatic chemicals, notably benzene and phenol, from petroleum. Some types of petroleum contain aromatic compounds anyway, but this source was inadequate to meet the growing demand. New ways of making aromatic chemicals from the aliphatic hydrocarbons, such as "platforming," had to be developed. This need to introduce new processes and entirely new types of chemical plant led to the eclipse of the industrial organic chemist by the chemical engineer.

Most historians of the chemical industry have assumed that the changeover to petrochemicals was more or less inevitable, if only because petroleum was so cheap in the late 1950s and the 1960s. This consensus has been challenged by Raymond Stokes, who has argued that "the fate of coal chemistry was far from evident; even in the latter half of the 1950s it continued to dominate worldwide production methods and to provide apparently viable alternatives to petrochemicals." While his analysis of the German industry has been a useful counterbalance to the technological and economic determinism implicit in earlier studies, it is indeed hard to see how it could have turned out otherwise, given the economic and political power of the U.S. and its largest corporations. Other countries and other firms could have followed a different path, as technologically alternative routes were available—notably fermentation-based ethanol and coal-based acetylene—but these options could only survive in countries where politics rather than industrial economics were paramount, such as East Germany or South Africa.

The Oil Crisis and its Impact

The oil crisis of 1973, when the Arab-dominated Organization of Petroleum Exporting Countries

(OPEC) quadrupled oil prices was a major shock to the organic chemical industry. Higher oil prices destroyed the boom in petrochemicals, which had seen increasing profits during the 1960s, and the established players were fearful that the oil-producing countries would develop their own petrochemical plants. For the remainder of the 1970s—with another "oil shock" in 1979—there were fears that petrochemicals might become uneconomical. The industry took steps to reduce its energy consumption and to increase production yields, factors that were low priorities while oil was cheap. Perhaps the greatest effect of the oil crisis was the acquisition of the Conoco Oil Company by the chemical giant DuPont in 1981. DuPont came to regret its move into the petroleum industry, and it divested the oil business in 1998.

Having weathered the oil crises, the petrochemical industry saw profit levels decline to historically low levels amid increasing public concerns about the industry's impact on the environment. Some companies have continued their historical links with petrochemicals, including Dow and BASF. Others have turned to low-tonnage but high-margin specialty chemicals and biotechnology. In the 1990s, even oil companies have moved to divest their petrochemical interests. To a certain extent, private entrepreneurs—notably Jon Huntsman—filled the gap by acquiring petrochemical units sold off by major concerns.

The Future

Although petroleum and natural gas supplies are fairly secure at present, governments are increasingly imposing statutory requirements on industry to boost the share of renewable energy and, almost in passing, promote renewable chemicals. At present, this sector is very small and has focused on fermentation (mainly ethanol), cellulose modification, and bioplastics. Perhaps surprisingly, wood distillation has not been prominent, but there has been a revival of interest in Bergius's process for the production of sugars from wood. Nonetheless, petrochemicals will continue to dominate the organic chemical industry for the first two or three decades of the twenty-first century.

Importance of Feedstocks

There are two ways of looking at feedstocks. The first is to have an end product in mind and then choose the appropriate raw material. Alternatively, a raw material may be cheap and abundant, even a waste product, and new uses can be developed for it. The chemical industry has been brilliant at mediating between the two approaches. It has taken waste materials and converted them into valuable products. It has perceived needs and created new ways of meeting demands. It has also mediated in another important way, namely between abundant (and hence cheap) raw materials, perhaps even the byproduct of another process, and the desired products. The result of this complex mediation is often a compromise. The raw material may be more scarce or more expensive than is desirable, and the industrial chemist may look for a more suitable raw material. On the other hand, the chemist may retain the original starting material but search for a reaction with a higher yield or one with a better end result. Although a particular raw material may have been taken up because it was originally a waste product or in surplus, its success can result in the need to manufacture it especially for that process, thereby losing the rationale for its initial selection. In this way feedstocks promote innovation. Any attempt to modernize or expand the industry usually means changing feedstocks. At the same time, however, there is usually a strong element of continuity in the processes used. The history of the organic chemical industry is one of evolution rather than revolution. Modern petrochemistry owes a large debt to the development of acetylene chemistry and the development of coal-to-oil processes in the first half of the twentieth century. A renewable chemical industry will be at least partly based on technologies that were introduced in the same period.

See also **Biotechnology; Chemicals; Dyes; Fertilizers; Oil from Coal Process; Nitrogen Fixing; Plastic, Thermoplastics; Reppe Chemistry; Synthetic Rubbers; Solvents**

Peter Morris

Further Reading

Aftalion, F. *A History of the International Chemical Industry From the "Early Days" to 2000*, 2nd edn. Translated by Benfey, O.T., Chemical Heritage Foundation, Philadelphia, 2001.

Greenaway, F. *et al*. The chemical industry, in *A History of Technology*, vol. VI, *The Twentieth Century, c.1900–1950*, Part I, Williams, T.I., Ed. Clarendon Press, Oxford, 1978, pp. 514–569.

Haber, L.F. The chemical industry, a general survey, in *A History of Technology*, vol. VI, *The Twentieth Century, c.1900–1950*, Part I, Williams, T.I., Ed. Clarendon Press, Oxford, 1978, pp. 499–513.

Herbert, V. and Bisio, A. *Synthetic Rubber: A Project That Had to Succeed.* Greenwood Press, Westport, CT, 1985.

Marcinowsky, A. and Keller, G. Ethylene and its derivatives: Their chemical engineering and genesis at Union

Carbide Corporation, in *A Century of Chemical Engineering*, Furter, W., Ed. Plenum Press, New York, 293–352.
Morris, P.J.T. Ambros, Reppe and the emergence of heavy organic chemicals in Germany, in *Determinants in the Evolution of the European Chemical Industry, 1900–1939: New Technologies, Political Frameworks, Markets and Companies*, Travis, A.S. et al., Eds. Kluwer, Dordrecht, 1998, pp. 89–122.
Reuben, B. G. and Burstall, M. L. *The Chemical Economy: A Guide to Technology and Economics of the Chemical Industry*. Longman, London, 1973.
Russell, C.A., Ed. *Chemistry, Society and the Environment: A New History of the British Chemical Industry*. Royal Society of Chemistry, Cambridge, 2000.
Spitz, P.H. *Petrochemicals: The Rise of an Industry*. Wiley, New York, 1988.
Stokes, R.G. *Opting for Oil: The Political Economy of Technological Change in the West German Chemical Industry, 1945–1961*. Cambridge University Press, Cambridge, 1994.

Fertility, Human

Assisted human reproduction technologies (ARTs) have flourished in the mid- to late-twentieth century. Greater understanding of human biology and reproduction has led to technological developments to assist individuals or couples experiencing infertility due to a wide range of indications. The field has evolved from a combination of technological advances (such as laparoscopy and transvaginal ultrasonography) together with pharmaceutical developments (notably purified extracts of human menopausal gonadotropic hormones) and theoretical knowledge and practical techniques taken from gynecology, genetics, urology, and associated medical specialties. However, the origins of many current ART techniques can be traced to early practices in animal husbandry.

Humans have used some less technological methods to promote their reproduction without the involvement of medical professionals, notably artificial insemination by donor (AID) or by husband/partner (AIH). The introduction of semen or concentrated specimens of spermatozoa (sperm) into a woman's reproductive tract by noncoital means can be successfully performed with instruments as simple as a turkey baster. However in recent years, fears about donor health status, risk of infection (HIV and otherwise), and legal issues (such as establishing paternity) have caused most AI to be performed in medical clinics under a physician's supervision. Some doctors avoided paternity issues by mixing sperm from several donors including the male partner, but recent advances in genetic technologies allowing paternity testing using DNA have resulted in clarification in many jurisdictions of the legal standing of children born from AI (though issues remain for instance with custody and adoption of AID children born to lesbian couples). AI is sometimes coupled with use of gonadotrophic hormones to stimulate ovulation at the time of insemination to maximize the chances of fertilization occurring, although these drugs are associated with some risks to the women to whom they are given.

In vitro fertilization and transcervical embryo transfer (IVF-ET) had its first successful birth in 1978, in the clinic of Patrick Steptoe and Robert Edwards, who drew on embryological studies done for over 20 years in mice, rabbits, and other animals. The procedure involved laparoscopic aspiration of an oocyte (egg cell or more generally "egg") during a natural cycle (thus circumventing damaged fallopian tubes), followed by IVF using ejaculated sperm and transfer of the cleaving embryo into the woman's uterus. More generally in IVF-ET, eggs are harvested and mixed in Petri dishes either with donor sperm (AID), or with sperm from the male partner (AIH, if primary male infertility is not thought to be at issue), typically using the healthiest sperm to facilitate fertilization. Eggs may be obtained from the female or donated by another woman (e.g., in cases of premature ovarian failure, genetic abnormalities, or reduced egg production due to advanced maternal age). Most women undergo controlled ovarian hyperstimulation (as described above) prior to aspiration of eggs to increase the number of eggs that are viable. Surrogacy (the establishment of pregnancy in another woman who either donates an egg to combine with the male partner's sperm or carries a fetus produced through combination of the couple's own gametes) has occurred in cases where the female partner in a couple wishing to have a genetically-related child cannot carry a pregnancy (e.g., due to lack of uterus for congenital reasons or following hysterectomy), but the practice has been curtailed in recent years due to legal restrictions following custody disputes between surrogates and couples.

The number of fertilized embryos created and transferred differs according to anticipated success, typically related to the putative cause of infertility in the couple as well as the clinic's experience. But in recent years, improved methods have created higher success rates both in terms of creation of viable embryos as well as implantation of the embryo posttransfer (the latter had been and remains the major technological barrier to successful pregnancies via IVF). The result has been

multiple gestations (often resulting in subsequent "selective reduction"; that is, termination of one or more fetuses to avoid the increased risks associated with multiple births), as well as emerging social, ethical, and legal issues associated with the status and disposition of supernumerary embryos. Consequently, many clinics have adopted more conservative approaches to the number of embryos created and transferred at any one time, and there is legislation or guidelines in some places to limit the number of embryos that can be transferred during one IVF cycle. Supernumerary embryos can be cryopreserved (typically at the four- to eight-cell stage) for later IVF cycles, donation to other infertile couples, or under certain circumstances for research. By the mid-1980s, techniques for cryopreservation were sufficiently developed to allow successful pregnancies using ET with frozen embryos, which permitted women to avoid multiple cycles of ovarian stimulation. More recently, IVF-ET has been combined with preimplantation genetic diagnosis (PGD) techniques to allow testing of embryos for genetic diseases (the technique was originally developed as an alternative to prenatal diagnosis for fertile couples with known genetic risks) and chromosomal abnormalities to allow selection and transfer only of unaffected embryos.

A number of additional ARTs have been developed in the last 20 years. Gametic intrafallopian transfer (GIFT), which involves placement of eggs (which have been removed from the follicles) together with sperm directly into the oviducts for fertilization, was first described in 1985, and is used with women with fallopian tube problems. This technique quickly became very popular because it did not require sophisticated IVF culture systems and could be done in clinics with less ART expertise and without a full IVF laboratory, and because it produced better results than IVF, perhaps because fertilization occurs in a natural environment (greater success rates are also due to patient selection, as was later recognized). Zygote intrafallopian transfer (ZIFT) involves transfer of the zygote (a fertilized egg that has not yet divided) into the oviduct after IVF, but is less frequently used. Intracytoplasmic sperm injection (ICSI) is a popular micromanipulation technique used to enhance fertilization rates, particularly for men with a reduced sperm count or with impaired sperm motility, banked sperm (obtained prior to chemotherapy or radiation), or sperm obtained through electroejaculation (e.g., in those with spinal cord injuries or recently after death, the latter being ethically and legally problematic). Pregnancy can be achieved with only a single spermatozoon injected directly into the cytoplasm of the oocyte. The technique also can be combined with those allowing separation of male and female sperm to avoid birth of children of a particular sex (for reasons of sex selection, which is considered ethically controversial by many, or avoidance of sex-linked diseases).

See also **Artificial Insemination and *In Vitro*** Fertilization; Cloning, Testing and Treatment Methods; Genetic Screening and Testing; Genetic Engineering, Methods

RACHEL A. ANKENY

Further Reading

Chen, S.H. and Wallach, E.E. Five decades of progress in management of the infertile couple. *Fertil. Steril.*, 62, 665–685, 1994.

Cohen, C.B., Ed. *New Ways of Making Babies: The Case of Egg Donation*. Indiana University Press, Bloomington, 1996. Commissioned by the National Advisory Board on Ethics in Reproduction.

Edwards, R.G. and Steptoe, P. *A Matter of Life: The Story of a Medical Breakthrough*. Hutchinson, London, 1980.

Hildt, E. and Mieth, D., Ed. *In Vitro Fertilisation in the 1990s*. Ashgate, Aldershot, 1998.

U.S. National Research Council. *Medically Assisted Conception: An Agenda for Research*. Report of a Study by a Committee of the Institute of Medicine. National Academy Press, Washington D.C., 1989.

U.S. Congress Office of Technology Assessment. *Artificial Insemination: Practice in the United States: Summary of a 1987 Survey—Background Paper*. U.S. Government Printing Office, Washington D.C., 1988.

Royal Commission on New Reproductive Technologies. *Proceed with Care*. Minister of Government Services, Ottawa, 1993.

Seibel, M.M. and Crockin, S.L., Eds. *Family Building Through Egg and Sperm Donation: Medical, Legal, and Ethical Issues*. Jones & Bartlett, Boston, 1996.

Seibel, M.M., Bernstein, J., Kiessling, A.A. and Levin, S.R., Eds. *Technology and Infertility: Clinical, Psychosocial, Legal and Ethical Aspects*. Springer Verlag, New York, 1993.

Talbert, L.M. The assisted reproductive technologies: an historical overview. *Arch. Pathol. Lab. Med.*, 116, 320–322, 1992.

Warnock, M. *A Question of Life: The Warnock Report on Human Fertilisation and Embryology*. Blackwell, Oxford, 1985.

Wikler, D. and Wikler, N.J. Turkey-baster babies: The demedicalization of artificial insemination. *Milbank Quarterly*, 695–640, 1991.

Law and Ethics of A.I.D. and Embryo Transfer. Proceedings of a Symposium on Legal and Other Aspects of Artificial Insemination by Donor (A.I.D.) and Embryo Transfer, London, 1972. Ciba Foundation Symposium, vol. 17, Elsevier, Amsterdam, 1973.

Fertilizers

As the twentieth century opened, fertilizers were a prominent concern for farmers, industrialists, scientists, and political leaders. In 1898, British scientist William Crookes delivered a powerful and widely reported speech that warned of a looming "famine" of nitrogenous fertilizers. According to Crookes, rising populations, increased demand for soil-depleting grain products, and the looming exhaustion of sodium nitrate beds in Chile threatened Britain and "all civilized nations" with imminent mass starvation and collapse. Yet Crookes also predicted that chemists would manage to discover new artificial fertilizers to replace natural and organic supplies, a prophecy that turned out to encapsulate the actual history of fertilizers in the twentieth century.

Three basic nutrients—nitrogen, phosphorus and potassium (typically identified as NPK)—as well as several minor and trace minerals are essential requirements for productive agriculture. But nitrogen lay at the crux of fertilizer problem. The obvious solution lay in the atmosphere, where nitrogen forms 78 percent of the atmosphere's volume. In this context, German investigators led by Carl Bosch and Fritz Haber developed the process that became standard in industry—the synthetic production of ammonia by subjecting atmospheric nitrogen and hydrogen to high temperatures and pressures. German industrial giant BASF opened a nitrogen fixation plant in 1913, although the real impact of nitrogen fixation upon agricultural practice remained roughly two decades away. In the meantime, significant geopolitical battles continued over the price and supply of natural nitrogenous fertilizers, particularly the sodium nitrates of Chile. Chile's share of world nitrate fertilizer production fell from about 70 percent at the beginning of the century to under 2 percent at the end. Also, near the turn of the century, scientists demonstrated the connection between soil microbes and the ability of legumes to fix atmospheric nitrogen. This discovery created a burgeoning commercial market for legume inoculation bacteria, products that promised a new industry for "germ fertilizers" and "green manures."

Important changes affected other branches of the fertilizer industry as well. Since the early industrial era, farmers applied phosphate fertilizers derived from natural sources such as bones and phosphate rock. Since the middle of the nineteenth century, farmers typically applied phosphates in the form of superphosphates, produced by treating phosphoric minerals with sulfuric acid. Production of superphosphates peaked in 1952. Since World War II, however, ammonium phosphates have emerged as the more predominant form. Discoveries of rock phosphates in Florida and the Rocky Mountain states of the U.S. in the 1870s, in Algeria and Tunisia in the late nineteenth century, Morocco, the Kola Peninsula of the USSR, South Africa, and elsewhere have ensured an adequate supply for well-capitalized agriculturists up to the present.

In the preindustrial era, farmers applied potassium fertilizers in the form of wood ash and other natural byproducts. The potash industry was transformed in the late nineteenth century, particularly due to discovery of the Stassfurt salts of central Germany. Potash became an important political issue in World War I, as the U.S. abruptly cancelled its contracts with Germany without an adequate means of replacing them. An intense search for domestic supplies continued until potash deposits were discovered in New Mexico in the 1920s. The U.S. was a leading potash producer until of the middle of the twentieth century. Since then, the principal producers have been Canada and Russia, and additional reserves have been found in China, Thailand, Brazil, and elsewhere.

The fertilizer industry experienced further expansion after World War II due to improved techniques of producing granular, homogenized, and well-mixed fertilizers. These replaced the powder form, since they were easier to store and handle, and apply. Blended granular fertilizers were introduced in the U.S. in the 1950s. Manufacturers produced a wide range of NPK combinations appropriate for specific local soil and cropping conditions. In the 1990s, "precision farming" techniques have integrated technologies of soil nutrient sensors, global positioning satellites, and the machinery that applies fertilizer. As a result, well-equipped farmers continually and automatically adjust fertilizer delivery formulas as they drive down their fields.

Government policies also facilitated the rapid expansion of fertilizer consumption in the late twentieth century. Britain's 1947 Agricultural Act was a typical example, offering incentives for fertilizer use as part of a strategy to intensify agricultural production. In the U.S., the quasi-governmental Tennessee Valley Authority has been at the center of many technical innovations in fertilizer processing. In the former German Democratic Republic, state support for fertilizer manufactories exemplified the technocratic assumption that agricultural chemicals are an

inherent good. Yet the most rapid expansion has occurred in the People's Republic of China, where officials linked shortages of nitrogenous fertilizers with the famines of the late 1950s and early 1960s. Thus China invested heavily in nitrogen fixation technology, and emerged rapidly as the largest consumer of manufactured fertilizers, both in terms of total consumption and rates of application. This in turn has spurred the most rapid rise of human population of all time. Other Asian nations, including the Republic of China, the Republic of Korea, and India also have become major producers and consumers of artificial fertilizers.

In addition to obvious links to increased agricultural production, the modern fertilizer industry has been linked with a number of concerns beyond the farm. For example, the short-lived phosphate boom on the Pacific island of Nauru offers a telling case study of the social consequences and environmental devastation than can accompany extractive industries. Further, much of the nitrogen applied to soils does not reach farm plants; nitrates can infiltrate water supplies in ways that directly threaten human health, or indirectly do so by fostering the growth of bacteria that can choke off natural nutrient cycles. To combat such threats, the European Union Common Agricultural Policy includes restrictions on nitrogen applications, and several nations now offer tax incentives to farmers who employ alternative agricultural schemes. Nevertheless, the rapidly growing global population and its demand for inexpensive food means that artificial fertilizer inputs are likely to continue to increase.

See also **Nitrogen Fixation**

MARK R. FINLAY

Further Reading

Haber, L.F. *The Chemical Industry, 1900–1930*. Clarendon Press, Oxford, 1971.

Monteón, M. *Chile in the Nitrate Era: The Evolution of Economic Dependence, 1880–1930*. University of Wisconsin Press, Madison, 1982.

Nelson, L.B., *History of the U.S. Fertilizer Industry*. Tennessee Valley Authority, Muscle Shoals, Alabama, 1990.

Smil, V. *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production*. MIT Press, Cambridge, MA, 2001.

Welte, E. Die Bedeutung der Mineralischen Düngung und die Düngemittelindustrie in den letzten 100 Jahren. *Technikgeschichte*, 35, 37–55, 1968.

Wines, R.A. *Fertilizer in America: From Waste Recycling to Resource Exploitation*. Temple University Press, Philadelphia, 1985.

Fibers, Synthetic and Semi-Synthetic

Silk has long been prized as a fiber for fashion, and Robert Hooke proposed artificial silk at least as early as 1665. The invention of the electric light bulb in 1879 created a new incentive for an improved fiber: a very uniform carbon fiber was required for lamp filaments. In 1883, Joseph W. Swan invented a process of squirting a nitrocellulose solution through a dye into a coagulating bath of water and alcohol. A denitration technique was used to make the fiber less flammable. In 1885, he exhibited the material at the Inventions Exhibition as "artificial silk." Although Swan was the first to spin fibers from a nitrocellulose solution, the first commercial production of artificial silk took place in France in 1884 by Louis de Chardonnet. De Chardonnet's product, however, failed commercially as he did not denitrate the fibers to reduce their flammability, with disastrous results. In the first quarter of the twentieth century, the Chardonnet process gradually disappeared.

In Germany, Max Fremery and Johan Urban produced good quality fibers by using solutions of cellulose in a copper oxide and ammonia mixture. In 1899, production started in the Vereinigte Glanzstoff Fabriken AG. In England in 1884, Charles Cross and Edward Bevan discovered that alkali–cellulose could be converted into cellulose–xanthogenate, which could be dissolved in sodium hydroxide. This solution was called "viscose," from which fibers could be pulled. Together with Clayton Beadle, they established the Viscose Syndicate Ltd. at Kew. Viscose fibers were of very good quality, and gradually the viscose process became the main process for producing artificial silk.

After World War I there was no use for vast amounts of cellulose acetate lacquers that had been used for doping fabric-covered aircraft. Henry and Camille Dreyfuss used these lacquers to produce fibers. In 1921, they introduced acetate fibers to the U.S. "Acetate" made great inroads on the market.

After World War I it was very important to show that cellulose fibers were not merely a "surrogate" for silk, and the name "rayon'" was introduced in 1924 and gradually accepted worldwide. Though rising hemlines created a demand for cheaper alternatives to silk stockings, rayon "bagged" at the ankles and in the 1930s and 1940s uses turned more to industrial products. One of the main improvements was a "hot stretch" process to increase the strength of rayon. In this way rayon could replace cotton in tire cords. Although the basic features of the three processes

to produce rayon fibers (viscose—by far the most significant process by volume—acetate, and cuprammonium) did not change much, efficiency of production continued to increase. Between 1930 and 1940, DuPont's manufacturing costs of a pound of viscose fiber fell nearly 60 percent. Until the recession of 1929, the rayon industry was a golden business.

However, the basic science of rayon was not understood. Chemists explained the characteristics of these materials by postulating forces between circular molecular structures. Most important in challenging these established concepts was Hermann Staudinger. He claimed in 1920 that normal molecular bonds could explain the products of polymerization reactions. Compounds, which precipitated during these reactions, could be explained by assuming that hundreds of molecules merged into "macromolecules." The work of Wallace H. Carothers, at E.I. DuPont de Nemours, was crucial for the acceptance of Staudinger's concepts. From 1928 onward, Carothers proved the existence of macromolecules by synthesizing new, long-chain molecules. Julian Hill, one of his research group members, discovered that he could draw fibers from the melt of these materials. In 1934, Carothers's group succeeded in making a good-quality polyamide 6,6 fiber. Technological problems were enormous: the process for spinning the fibers differed greatly from any conventional process. At the end of 1938, DuPont launched the fiber as Nylon. Nylon stockings were an overwhelming commercial success until it was rationed to replace silk in parachutes and to reinforce airplane tires. In Germany, Paul Schlack had succeeded in producing a Nylon-like fiber by the end of the 1930s. The fiber was often called Nylon 6 as opposed to DuPont's Nylon 6,6. After the war, the nylon 6 patents of IG Farben were confiscated. Many corporations started Nylon 6 production, as the technology was freely available.

During the war, research on new fibers continued. The most important of the new fibers were acrylic fibers and polyester fibers, including terylene discovered in Britain by John R. Whinfield and James T. Dickson in 1941. ICI and DuPont started producing polyester fiber (terylene and dacron) in the early 1950s. Polyester fiber had very good mechanical properties and a better light stability than nylon. Because of its very good resilience, the main application of this fiber was initially in "wash and wear" textiles.

In 1942, IG Farben found a suitable solvent to spin acrylic fibers, and after World War II, they were commercialized by Bayer, DuPont and Monsanto. Acrylic fibers had comparable properties to wool, and the chemical, thermal and light stability of the fiber was rather good. Acrylic fibers were spun by a new spinning process: after spinning, the solvent evaporated when the fibers were led through a gas stream in a spin tower. This process is called dry spinning. Nylon and polyester were spun by a melt spinning process, which implies that the polymers are molten before spinning, and not dissolved. Rayon was produced by wet spinning, which implies that the polymer is dissolved. However, the solvent is later removed from the fiber in a bath.

Polypropylene fiber was developed at the end of the 1950s, based on a breakthrough in polymer science at the beginning of the 1950s in which ordered polyethylene and polypropylene polymers

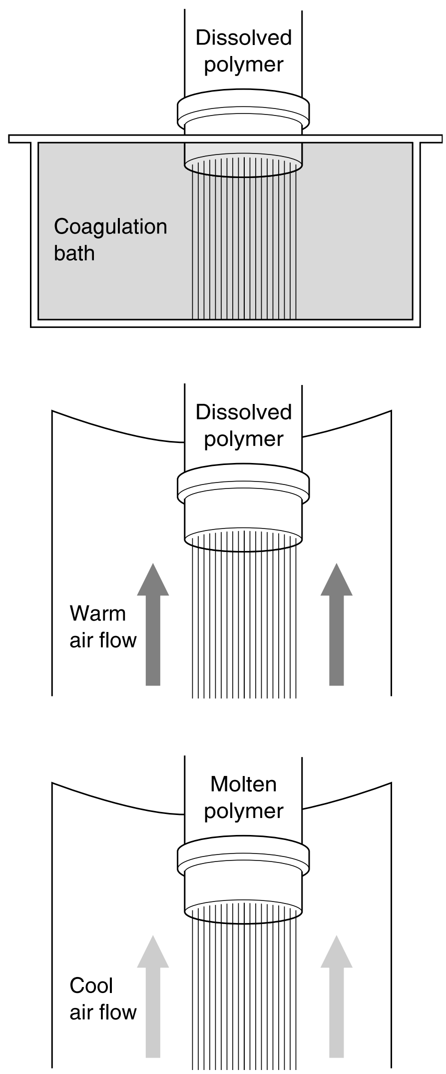


Figure 2. Wet spinning, dry spinning, melt spinning.

could be made catalytically. Polypropylene fiber was excellent for ship ropes, as its strength was high and it floated. Polyethylene was also used for fibers after DuPont invented the "flash spinning" technique by which very cheap, but low-quality polyethylene fibers could be spun. Polyethylene fibers were used to form a sheet-like wrapping material.

Particularly in the 1950s and 1960s, synthetic fibers conquered large shares of the market. Production efficiency and fiber properties increased considerably. For example polyester-spinning speeds increased from 500 meters per minute in 1961 to 4000 meters per minute in 1975, allowing fiber prices to drop dramatically in these years. Improved quality opened new markets. For example, nylon was good enough to be used in car tires after World War II, although tires reinforced with nylon were subject to so-called "flat spotting"; they had a flat spot when the car had been parked for a while.

At the beginning of the 1950s, scientists at DuPont achieved a breakthrough in polymerization. Their new method did not involve heating of intermediates. A lot of new polymers were made which could not be made before because the intermediates could not withstand heating. By this method, DuPont scientists developed the spandex fiber Lycra®. A high temperature-resistant fiber, Nomex®, was made based on a wholly aromatic polyamide. Similar fibers were introduced in Japan as Conex® and in the USSR as Fenilon®.

Other companies also started researching DuPont's low temperature polycondensation method. Celanese developed very heat-resistant and comfortable polybenzimidazole fibers. Protective clothing became more important when three of the Apollo I astronauts burned in their spacecraft in an accident during launching in January 1967.

The 1950s and 1960s were the glorious days of the fiber industry. Growth rates and profits were high. However, the market was sensitive to cyclical fluctuations, which sometimes caused financial problems at the fiber corporations. In 1967, the U.S. market collapsed. In 1970, the European fiber market was in crisis too. The recession that followed the oil crisis of 1973 put the fiber industry in the doldrums. The situation, especially in Western Europe, was dramatic. Many plants were closed down and some traditional fiber producers completely terminated their business.

In the 1960s, high-strength/high-modulus (HS/HM) carbon fibers were under development at various laboratories. These fibers were aimed at making composites. Carbon fibers were very good but brittle and expensive. They were therefore only used for military applications and in the aerospace industry. At several laboratories, it was found that HS/HM fibers could be made using aromatic or heterocyclic polyamides. In 1964, DuPont made a fiber that had an unusually high strength and modulus. Soon it was recognized that these higher strengths could possibly be used to make improved tire cords and a cheaper alternative for carbon fiber. In February 1970, DuPont announced its HS/HM fiber in public. The fiber was called Kevlar®. Other corporations had started researching HS/HM fibers too. Monsanto, Bayer and ICI made several contributions to the technology. AKZO, which improved DuPont's production process, got into an enormous patent litigation case with DuPont in the 1980s when it introduced Twaron®, a similar fiber to Kevlar. The Japanese company Teijin also developed an HS/HM fiber Technora®. In the Soviet Union, an HS/HM fiber was developed further for military use.

Kevlar was not as successful as DuPont had hoped. It was only used as a tire cord in very specialized tires. The U.S. tire cord market was gradually conquered by the steel wire reinforced radial tire, which had previously conquered Europe. New applications of Kevlar were ballistic protection, composites and replacement of asbestos.

In the 1980s, some corporations carried out research and development to develop fibers of even higher modulus and strength than Kevlar. Stanford Research Institute synthesized two polymers, PBO (poly-para-phenylene-benzo-bis-oxazole) and PBT (poly-para-phenylene-benzo-bis-thiazole), which were further developed by Dow Chemical and Toyobo. Akzo Nobel developed a similar fiber called M5®.

The polymers that were used for HS/HM fibers had a rigid rod structure. However, one also succeeded in making HS/HM fibers from flexible polymers: In 1979, scientists of the Dutch corporation DSM succeeded in spinning HS/HM fibers from a polyethylene gel and stretching it. By this process, they created an ordered structure in the polymers. This polyethylene fiber was in some applications a competitor for HS/HM aramid fibers because it was even stronger. Due to its low melting point it could not compete with aramid fibers in applications in which higher temperatures played a role.

See also **Composite Materials; Plastics, Thermoplastics; Synthetic Resins**

KAREL MULDER

Further Reading

Beer, E.J. *The Beginning of Rayon*. Beer, Paignton, Devon, 1962.

Coleman, D.C. *Courtaulds: An Economic and Social History II: Rayon*. Clarendon Press, Oxford, 1969.

Farber, E. *Great Chemists*. Interscience Publishers, New York, 1961.

Hermes, M.E. *Enough for One Lifetime: Wallace Carothers, Inventor of Nylon*. American Chemical Society/Chemical Heritage Foundation, Washington, 1996.

Hollander, S. *The Sources of Increased Efficiency: A Study of DuPont Rayon Plants*. MIT Press, Cambridge, MA, 1965.

Hongu, T., Phillips, G.O. *New Fibers*. Horwood, New York, 1990.

Klare, H. *Geschichte der Chemiefaserforschung*, Akademie Verlag, Berlin, 1985.

Morawetz, H. *Polymers: The Origins and Growth of a Science*. Wiley, New York, 1985.

Morris, P.J.T. *Polymer Pioneers*. Beckman Center for History of Chemistry, Philadelphia, 1986.

Mulder, K.F. Choosing the corporate future: Technology networks and choice concerning the creation of high performance fiber technology. University of Groningen, dissertation, 1992.

Mulder, K.F. A battle of giants: the multiplicity of industrial R&D that produced high-strength aramid fibers. *Technol. Soc.*, 21, 37–61, 1999.

Seymour, R.B. and Porter, R.S., Ed. *Manmade Fibers: Their Origin and Development*. Elsevier, London, 1993.

Staudinger, H. *Arbeitserinnerungen*. Dr. Alfred Huthig Verlag GmbH, Heidelberg, 1961.

Film and Cinema, Early Sound Films

Silent films were rarely silent. Scarcely had the gasps and screams at the Lumière brothers' train arriving at La Ciotat station died down before its successors were being screened to music provided by anything from a piano to a 100-piece orchestra. The technology needed to provide films with a "tailored" sound accompaniment had existed for almost 20 years, and in September 1896 Oskar Messter showed "sound films" in Berlin using synchronized Berliner disks. Many others followed, including talking films, included in a program at the Paris Exposition by Clement Maurice on 8 June 1900.

These were able demonstrations that pushed existing film and sound technology boundaries, but producing a reliable sound film system required three major advances:

- Providing a foolproof method of picture/sound synchronization
- Recording a wider sound frequency range at greater levels than existing acoustic methods
- Reproducing these loud and clear enough to fill a cinema auditorium

Ingenious methods of synchronizing sound recordings with films were devised, but all suffered from the fragility of records over repeated reproduction and during transport, and the risk of picture and sound going "out of sync." The solution lay in combining the two together on film—a process known as "sound-on-film." Frenchman Eugene Lauste patented the first such system in London on 11 August 1906, but it took seven years to realize this, when war intervened. Lauste's "sound gate" used two slotted iron grids through which light passed to the film. One grid was fixed, while the other slid up and down over it in response to a signal from a microphone relayed through an electromagnet, resulting in a variable density soundtrack.

With this impetus lost it fell to independent researchers to develop sound-on-film systems. Josef Engl, Hans Vogt, and Josef Engl devised Germany's "Tri-Ergon" process. Another variable density process, it used a photoelectric cell to turn sound into electric waves and then into light recorded photographically on to the film's edge. In the projector a photoelectric cell reconverted these into an electrical signal. Tri-Ergon was first demonstrated in Berlin on 17 September 1922.

The American "Phonofilm" process was developed by electronics and radio pioneer Lee de Forest in 1920. A valve produced a fluctuating light pattern in response to an electrical sound signal, exposing a series of light and dark areas on the side of a film. On projection this was read by a photocell and converted back to a sound signal. Phonofilm debuted in New York on 15 April 1923. De Forest's system enjoyed greater commercial success, and over 30 cinemas had been equipped to show its mainly musical short films within a year, but it failed, lacking major studio and distributor support.

With sound films possible technically, ways of improving sound recording and reproduction were needed. Stimulus to develop and improve sound recording came from U.S. telephone companies, notably American Bell, who formed American Telephone & Telegraph (AT&T) on 3 March 1885. A Bell subsidiary—Western Electric—became AT&T's research arm. In 1907 this was centered in New York, under director Theodore Vail, and between 1913 and 1926, breakthroughs were made there in electrical recording, playback, and amplification.

E.C. Wente patented the condenser microphone on 20 December 1916. It translated sound waves into electrical ones that could be transmitted by the vacuum tube amplifier. Improved over the years, in

1926 it became the Western Electric 394-W microphone, used to produce the first true sound films.

By 1920 electrical recording had been developed by Henry C. Harrison. It used Wente's microphone, a tube amplifier, a balanced-armature speaker, and a rubber-line recorder, recording sound in the range 50 to 6,000 Hz, an improvement over the 250 to 2,500 Hz range of acoustic recordings.

Harold Arnold developed the tube amplifier in April 1913. Based on Lee de Forest's Audion tube, Arnold found that electron flow across its electrodes improved if they were in a vacuum. Arnold built his first vacuum tube on 18 October 1913.

C.W. Rice and E.W. Kellogg, and others developed the moving coil loudspeaker. Henry Egerton patented the first balanced-armature loudspeaker driver on 8 January 1918, and E.C. Wente developed the moving coil speaker (patent filed 4 August 1926). He used a moving coil or diaphragm mechanism in a strong magnetic field. Designed to drive a theater speaker, it was installed at the Warner Theater, New York for the premiere of Warner's "Don Juan" in August 1926.

Despite sound film technology being ready by 1925, there was little interest from major studios, and it fell to an aspiring one—Warner Bros.—to seize the initiative. Sam Warner learned about Western Electric's achievements when Warner Bros. built a radio station. He saw its potential for providing a full orchestral accompaniment to Warner films wherever they were screened. On 20 April 1926 they formed the Vitaphone Corporation with Western Electric to develop this, but opted for a synchronous sound-on-disk system. A 406-mm (16-inch) disk, revolving at 33 1/3 rpm, played simultaneously with the film, synchronized by two motors held at the same speed by an electric gear. Interconnected by slip rings, the interchange of power between the armatures ensured correct synchronization when the film started, with the power source's frequency then maintaining this.

Many Vitaphone short films were made, but Warner gave its greatest showcase with the production and premiere of "Don Juan" starring John Barrymore, featuring a full-length orchestral soundtrack with sound effects. Its premiere, on 6 August 1926, left a bemused audience but an enthusiastic press.

By the end of 1926, Warner had produced 100 Vitaphone shorts, but the cost, and those of equipping cinemas for them, almost crippled the company. They had one last attempt, adapting Samson Raphaelson's play "The Jazz Singer," with Al Jolson reprising his stage role in the lead. Intended only to have synchronized music and singing, audience reaction at the 6 October 1927 premiere was greatest to Jolson apparently talking to them from the screen. Ironically, his first words—"Wait a minute. Wait a minute. You ain't heard nothin' yet! Wait a minute, I tell you. You ain't heard nothing. You wanna hear toot-toot-tootsie? All right. Hold on"—were largely ad-libbed. Ignorant of this, the audience stood and cheered. Despite only containing 354 spoken words, The Jazz Singer's triumph firmly established talking pictures.

By cruel fate, Sam Warner died of a brain hemorrhage the night before the premiere. His surviving brothers—Harry, Albert, and Jack—were at his bedside, also missing their moment of glory. In the longer term, use of the Warners' synchronous Vitaphone system faded in favor of a sound-on-film system called Movietone, developed by the rival Fox Film Corporation.

See also **Audio Recording, Mechanical; Audio Recording, Electronic Methods; Audio Systems; Film and Cinema: High Fidelity to Surround Sound; Loudspeakers and Earphones**

PAUL COLLINS

Further Reading

Guy, R. *Warner Bros: Fifty Years of Film, 1923–1973*. LesLee Productions, Los Angeles, 1973.

Hirschhorn, C. *The Warner Bros. Story*. Octopus Books, London, 1979.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, Cambridge, 1995. An excellent idea of the interrelatedness of innovations in sound recording, radio and cinema.

Robertson, P. *The Guinness Book of Film Facts & Feats*. Guinness Superlatives, Enfield, Middlesex, 1980. The basic facts of who did what when.

Toulet, E. *Cinema is 100 Years Old*. Thames & Hudson, London, 1995. Covers the impact of cinema, and of the many innovations in production and presentation tried in its first decade or so.

Film and Cinema: High Fidelity to Surround Sound

While motion picture soundtracks typically generate less excitement than the visual content, the sound recording devices used in the motion picture industry have consistently been at the cutting edge of technological advancement. From the 1920s to the 1950s, these technologies reached the public in theaters long before anything similar was available from broadcasts or for home use. In fact, many of the early innovations in high fidelity recording and

reproduction were created in the context of motion picture production and exhibition, while the phonograph, radio, and television lagged behind.

There were experimental linkings of sound and motion pictures from the earliest days of the cinema, and early designers of these systems struggled merely to provide sound at minimally acceptable levels of volume and quality. The nearly worldwide adoption of sound-on-film by the early 1930s corresponded to relatively favorable economic conditions in the motion picture industry, particularly in the U.S., and this encouraged even more experimentation with new audio techniques.

Some of the most notable achievements of the 1930s were related to stereophonic sound. Long before it was practical to introduce this technology to the public via the phonograph or radio, audiences in some places heard stereo soundtracks accompanying a small number of feature films. A landmark was *Fantasia*, the animated film by Walt Disney Studios (U.S.), which employed sound recorded on numerous separate optical tracks. These were then mixed down to three channels (left, right and center) of audio for exhibition along with a fourth "control" track, which was not audible but contained information that automatically controlled the volume of each of the three audio tracks. Only two theaters purchased the U.S.$85,000, 54-loudspeaker "Fantasound" system needed to reproduce these films on screen, but a traveling exhibition toured the U.S. when the film opened in 1940.

At the end of World War II, many motion picture producers adopted magnetic recording technology, which was known before the war but rarely used outside Germany. Magnetic recording was substituted for optical recording in the studios primarily because it was much less expensive to use; at a time when television was cutting deeply into theater attendance, cost cutting was imperative. However, that cost saving did not apply to the exhibition of films, and most theaters retained their optical-soundtrack projection equipment through the late 1980s. While studios repeatedly tried to introduce new theater systems using multichannel, high-fidelity sound, most exhibitors resisted. The theater was no longer at the forefront and many innovations in movie sound technology made after the 1950s were preceded by similar innovations in broadcasting or in home high-fidelity systems.

However, the experimental technologies of the 1950s are usually cited by film historians as great landmarks. One of the most notable examples was Cinerama, one of several widescreen formats that Hollywood studios believed would bring customers back to the theaters in the 1950s. Besides its remarkably wide screen, Cinerama featured a seven-track magnetic soundtrack, carried on a separate 35 mm film run on a player that was operated in parallel to multiple motion picture projectors. Like Fantasound before it, Cinerama was so expensive to exhibit that it saw only limited use. Somewhat more successful were systems based on a double-width, 70 mm film on a single projector. All of these used some variation of multichannel sound, and some used magnetic rather than optical soundtracks for theater reproduction. Perhaps the most commercially successful of these was Todd-AO (promoted by film producer Michael Todd and the American Optical Company), which used six audio channels.

A series of highly successful innovations was offered by Dolby Laboratories (U.K., later U.S.) beginning in 1965 with the introduction of what came to be known as Dolby A. This was a noise-reduction technology used to improve recordings made in the studio before they were released to theaters. Although used initially in the phonograph record industry, the first motion picture soundtrack made using Dolby A was *A Clockwork Orange*. Released in 1971, the movie was typical of the Dolby releases of the day in that it was originally recorded on multitrack magnetic recorders, mixed using Dolby noise reduction, but released in ordinary monophonic form, usually with an optical soundtrack. However, the next year Dolby introduced an improved optical soundtrack technology and the short film *A Quiet Revolution* was released to demonstrate to theater chain owners the value of using Dolby noise reduction equipment in the exhibition of these films. While this technology did not succeed, some theaters did begin to improve their audio equipment.

Sensurround, a multichannel system for theaters, was introduced as a sort of novelty with the film *Earthquake* in 1974. An optically recorded, inaudible control track triggered the reproduction of very low-frequency sounds, which were used to add emphasis to the soundtrack at key points (such as the rumbling of an earthquake). Sensurround-like systems would eventually evolve into the current "Surround Sound," but meanwhile the Dolby Laboratories once again introduced a new multitrack system. This one, called Dolby Stereo, electronically combined four soundtracks onto just two tracks for the final release print. The four tracks included left, right, and center channels plus a "surround" channel for special effects. Dolby Stereo could be reproduced by adding relatively

inexpensive accessories to existing projectors. The first release in Dolby Stereo was *A Star is Born* in 1976. Following the advent of Dolby Stereo it became more common to advertise a film's sound technology along with its cast. This helped generate a popular interest in film sound, and along with the consolidation of exhibition and production companies, the rate of adoption of new theater technologies began to accelerate.

Digital recording techniques were tried by filmmakers from the early 1980s, although they were not in widespread use until the 1990s. An early optical digital playback system introduced in 1990 by the Eastman Kodak Company (U.S.) was not as successful as Dolby Digital, introduced in 1992. In this new system, a digitized version of the soundtrack was placed in the tiny spaces between the sprocket holes on the exhibition copy of the film, thus leaving room for a conventional analog soundtrack at the edge to be used as a backup or in theaters that had only the standard projectors. Under various names, the software algorithms developed for Dolby Digital have also been adapted for other formats, such as home theater and DVD discs. With the introduction of digital recording and playback systems, motion picture producers and movie theaters are once again acting as the channels for the introduction of new audio technologies to the public.

See also **Audio Recording, Compact Disc; Audio Recording, Electronic Methods; Audio Recording, Mechanical; Audio Recording, Stereo and Surround Sound; Audio Recording, Tape; Audio Recording, Wire; Audio Systems; Loudspeakers and Earphones**

DAVID MORTON

Further reading

Fielding, R., Ed. *A Technological History of Motion Pictures and Television*, University of California Press, Berkeley, 1967.

Weis E. and Belton, J. *Film Sound: Theory and Practice*. Columbia University Press, New York, 1985.

Film and Cinema: Sets and Techniques

During the early years of silent cinema until around 1909, interiors for film sets resembled theatrical and vaudeville stages with painted backdrops and limited three-dimensional representation. Action was depicted at long-shot distance immediately in front of the backdrop creating a shallow playing area. The earliest narrative films consisted of a single shot of actions on a stage and spectators were positioned in relation to the action as though viewing a traditional theatrical presentation. After 1911 the development of the American film industry in Hollywood as one based primarily on considerations of narrative storytelling over the esthetics of form, meant that a preoccupation of early studio set technicians and cinematographers, designers, producers and artistic directors was a drive toward greater narrative progression and the realistic representation of action on screen. Innovations in editing through the use of the fast film stocks characteristic of silent cinema had two profound cinematic effects.

First, continuity editing, as it came to be known, meant that scenes could be broken up into discrete shots, creating a break from theatrical conventions and the creation of a truly cinematic space. The standard continuity technique is shot-reverse shot and crosscutting. The former, used in filming two-way conversation, works in conjunction with editing and camera angles in order to produce the effect of seamless movement between speaking characters. Shots are taken from one character's point of view or over the shoulder and intercut with those from the other person. Crosscutting, pioneered and refined in the work of D. W. Griffith, refers to the alteration of shots, this time between scenes. It implies multiple actions in different locations that are occurring simultaneously. Techniques of continuity editing, in hiding the constructed nature of the action represented on screen, produces a second cinematic effect contributing to the establishment of the particular "look" or esthetic of American film. The codes and practices of editing have remained largely consistent with these early innovations and have been little affected by subsequent developments in technology. This emphasis on narrative continuity led to the early identification of American film with realism. Cinematic realism holds that what is represented on screen accurately reproduces that part of the real world to which it refers. Thus, American film through the use of continuity editing techniques introduced a realism of cinematic representation and positioned film spectators close to the action. This reversed technique that led to the distancing of the audience from the spectacle, common to European formalism and experimental cinema, offering audiences the illusion that they are watching a seamlessly coherent and wholly realistic representation of reality.

The realistic codes of American cinema also demanded the representation of three-dimensionality, the cinema frame as a window on the world extending beyond the limits of the frame into space off-screen. To this end developments intended to

produce realistic cinematic effects occurred in liaison by cinematographers and art directors working with miniatures and techniques of composite photography such as glass shots and mattes. Glass shots worked to give cinematic depth through the technique of shooting through a clear pane of glass containing either a painted or photographic image. An early example of special effects, glass shots were used widely in the 1920s and 1930s and featured famously and extensively in the work of M. C. Cooper and E. B. Schoedsack in *King Kong* (1933).

By the 1920s, depth and three dimensionality was also being added to set design through modifications to stage architecture. The arrangements of props, increasing use of multiple room stage constructions, artificial lighting effects and careful construction of *mis-en-scéne* were all directed at creating the sensory illusion of depth and the participation of the audience in the screen space. Early attempts to represent the illusion of depth graphically were most convincing on location filming but proved incongruous when edited beside the shallow effects offered through traditional stage setting. However, developments as early as 1908 ushered forward techniques of deep focus cinematography through modifications to lens lengths and the use of wide angles, narrow apertures, and the manipulation of lighting, and provided for multiple planes of action area. As a technique of shooting, deep focus ensured that realism in depth could be achieved by allowing a number of actions to take place in multiple planes simultaneously during a single shot. Deep focus cinematography worked in combination with the development of techniques to enhance depth in set design. The illusion of depth offered through deep focus cinematography quickly became both the norm and the signature of American narrative cinema, though it was not fully realized until the pioneering work of Gregg Toland in the 1930s.

Toland, considered the greatest cinematographer of his age, experimented with arc lamps, wide angles, lens coating and fast film to produce a distinctive deep focus impression. The culmination of his work can be seen in *Citizen Kane* (1941) where Toland creates enormous and sumptuous spaces that achieve both great height as well as great depth. Toland's greatest innovation in *Citizen Kane* is considered to be his use of unusually long takes, using static cameras to achieve incredible shot depth of long duration. Toland's use of long takes added to the canon of deep focus techniques and was later developed in conjunction with the Garrett Brown's Steadicam to provide both an unusually long depth perception coupled with unerring rapid movement in *The Shining* (1980).

Contemporaneously, developments in special effects techniques from the pioneering work done in the 1920s and 1930s have become the modern signature of Hollywood productions. Common photographic techniques such as fades, wipes, and dissolves have been augmented by techniques such as rear projection in which live action is combined with painted backdrops or miniatures. Rear projection was used to fantastic effect by George Lucas in *Star Wars* (1977), the film considered to have reinvented the effects-led blockbuster film. Since then, digitization; that is, the use of electronically programmed motion control and computer graphics in which the fantastic is able to be represented realistically, has become ubiquitous in Hollywood productions in the wake of Steven Spielberg's breakthrough work of *Jurassic Park* in 1993.

Taken together, continuity editing, depth of field cinematography, composite special effects photography, increasingly sophisticated set design, and staging through mis-en-scéne, have each contributed to giving visual clarity and realism to multiple planes of action simultaneously. This has achieved the cinematic effect of greatly enhancing spectator perception of screen space by extending it forward toward the audience and approximating what the French theorist André Bazin, referring admiringly to American film production, called "total cinema."

See also **Cameras, Lens Designs: Wide Angle and Zoom; Film and Cinema, Early Sound Films; Film and Cinema, Widescreen Systems**

ANDREW PANAY

Further Reading

Allen, R.C., Gomery, D. *Film History: Theory and Practice*. McGraw-Hill, New York, 1985.
Bordwell, D., Staiger, J. and Thompson, K. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. Routledge, London, 1988.
Dick, B.F. *Anatomy of Film*. St. Martins Press, New York, 1998.
Katz, S.D. *Film Directing: Shot by Shot*. Michael Wiese Productions, Ann Arbor, MI, 1991.
Maltby, R. *Hollywood Cinema*. Blackwell, Oxford, 2000.
Nowel-Smith, G, Ed. *The Oxford History of World Cinema*. Oxford University Press, New York, 1996.

Film and Cinema: Widescreen Systems

The term "widescreen cinema" has been precisely defined by J. Belton as "a form of motion picture

production and exhibition in which the width of the projected image is greater than its height, generally by a factor of at least 1.66:1." This restricted definition of widescreen excludes other large-screen formats such as the Imax and Omnimax formats, which produce a width to height ratio (or "aspect ratio") of 1.43:1.

Although widescreen formats vary, they all share in common the stretching of screen width far beyond the initial standard set by the film industry of 1.33:1. This standard later came to be called "narrow-screen" in contrast to the new widescreen industry standard. The previous format was based on 35 mm film gauge with a negative image of 1 inch wide by 0.75 inches high, producing a 4:3 image aspect ratio of 4 units wide by 3 units high. Although some experiments were undertaken with widescreen in the 1920s, "narrow-screen" endured with relatively minor variations as the industry standard for 64 years between 1889 and 1953 and remained the standard twentieth-century format for the television screen, 16 mm films and 8 mm home movies.

As with most developments in cinematic technology widescreen experiments in the 1920s like Magnascope in the U.S. and Polyvision in Europe were dismissed as short-term novelty gimmicks of no lasting value. That is, until the development of Cinerama by Fred Waller transformed established attitudes to screen dimensions. Waller was a ceaseless innovator, experimenting in a number of fields. His independent studies of depth perception established that the convincing illusion of three-dimensionality depended on peripheral as well as binocular vision. He was able to combine this insight with technical knowledge garnered from his work with Paramount's Special Effects Department. His Cinerama camera set three lenses at 48-degree angles to each other to produce a composite angle of view of 146 degrees by 55 degrees, closely matching the angle of view of human vision at 165 degrees by 60 degrees. Projected onto a deep curved screen by a three-sided multicamera system and supplemented by stereo sound, Cinerama seemed to surround the viewer at every point of visibility and heightened the viewing sensation as one of complete immersion in hugely inflated but realistically depicted moving images of unprecedented depth. The curved screen solved for Waller the problem of fitting the field of peripheral vision within theatrical space: a flat screen would have needed to be the length of a city block.

Cinerama was launched to huge acclaim on 30 September 1952 with the travelogue film *This is Cinerama*. Cinerama failed to follow up on its initial success and was only adopted by a major studio, MGM, ten years later to make two feature films, *How the West Was Won* and *The Wonderful World of the Brothers Grimm*. It was limited by its own expensive, specialized technology, designed more to showcase exhibitions than for standard feature films in conventional theaters. Cinerama films were shot and projected at 26 frames per second, rather than the standard 24, and the three-strip Cinerama production format used three and a half times more negative film stock as conventional processes. Moreover, clearly visible vertical lines on the screen were formed by three separately projected triptych images, badly distorting the desired optical illusion of seamless reality. Not until 1963 did Cinerama resolve the problems caused by multicamera perspective by converting to Ultra Panavision, a widefilm technology that stored images and sound on a single strip of 65 mm negative film, which could be projected in 70 mm Cinerama theaters or printed down to 35 mm CinemaScope for conventional theaters. This solution was at the cost of the uniqueness of Cinerama's full glory and the ambition of its original aspect ratio. Even innovatory productions like *2001: A Space Odyssey* could not halt Cinerama's steady decline and ultimate demise in May 1978.

Cinerama's lasting contribution is in the stimulation it provided for other, less expensive and technically complex widescreen processes like CinemaScope (1953) and Todd-AO (1955). Unlike Cinerama, CinemaScope was the product of the research department of a major film studio—Twentieth Century Fox. On 2 February 1953, Fox announced that it would make all of its subsequent movies in CinemaScope. Only seven months later, on 16 September 1953, Fox issued its first widescreen release, *The Robe*.

CinemaScope's rapid development and successful release crucially depended on combining innovations in computerized lens design, acetate film stock, and Eastman Color film, as well as in television innovations in magnetic recording equipment and screen materials. Critical to this was Fox's rights to the anamorphic optical system of cylindrical lenses, the Hypergonar, developed in France in 1927 by Henri Chretien. Attached to a standard camera lens, the Hypergonar compressed a wide horizontal image by a factor of 2 onto standard 35 mm film during recording without affecting the vertical image. In projection a reverse anamorphic lens "'stretched' the image for widescreen display, permitting an extreme panoramic

image of 2.66:1. Fox's optical engineers worked with lens manufacturers Bausch & Lomb to modify Chretien's original square attachment to produce circular lens profiles, improving the overall consistency of projected light, resolution, depth of field and relative definition at the edges of the field.

With the screening of *Oklahoma!* on 10 October 1955, Fox's achievements in widescreen were eclipsed by Todd-AO, the process developed by Michael Todd and the engineering firm, American Optical. Todd-AO resolved the problem of magnification that beset other widescreen formats such as VistaVision and CinemaScope. By blowing up images to cover a screen 64 by 24 feet (19.5 by 17.3 meters), 35 mm film was magnified 330,000 times for CinemaScope, with some loss of image sharpness and resolution. Todd-AO recorded on 65 mm film and projected on 70 mm, the other 5 mm accommodated six magnetic soundtracks for stereo sound. The 65/70 mm format required a projected magnification of only 127,000 times to fill a curved 52 by 26 feet (15.8 by 7.9 meter) screen, ensuring sharper imagery. More expensive than 35 mm film, the 70 mm format acquired a prestigious reputation for cinematic quality. Todd assembled a talented team of researchers and contractors around American Optical, including the then current Director of the Institute of Optics, Brian O'Brien, to develop the process through a series of innovations. For example, a 128 degree "bug-eye" wide angle camera lens was developed and mounted on a single camera with three other lenses, 64, 48 and 37 degrees, to give a versatile range of coverage for long, medium and close-up shots.

In these few years in the 1950s widescreen developed out of past technical experiments and future expectations to meet market challenges and industry rivalries. It bequeathed a different sense of scale to cinema esthetics and provided an answer of sorts to the emergence of television as the leading mass medium. It lives on today in the form of 70 mm "blow-ups" where films such as *Star Wars* can be recorded on relatively cheap 35 mm stock and then "printed-up " to 70 mm for widescreen exhibition, shown in multiplex complexes with one or two 70 mm theaters. Widescreen's faint reflection can also be seen today in television's adoption of widescreen sets and video's use of "letterboxing."

See also **Cameras: Lens Designs, Wide Angle, and Zoom; Film and Cinema: High Fidelity to Surround Sound; Film and Cinema: Sets and Techniques; Television, Digital and High Definition Systems**

ALEX LAW

Further Reading

Bazin, A. Will CinemaScope save the Cinema? *Esprit*, 207–208, 672–683, 1953. Translated by Jones, C. and Neupert, N. and reprinted in *Velvet Light Trap, Review of Cinema*, 21, 1985. Bazin, A. *What is Cinema?* vol. 1. Editions du Cerf, Paris, 1958. Translated by Gray, H., University of California Press, Berkeley, 1967.

Belton, J. CinemaScope: the economics of technology. *Velvet Light Trap, Review of Cinema, Widescreen Issue*, 21, pp. 35–43, 1985.

Belton, J. *Widescreen Cinema*. Harvard University Press, Cambridge, MA, 1992.

Mitchell, R. History of widescreen formats. *American Cinematographer*, 68.5, 1987.

Spellberg, J. CinemaScope and ideology. *Velvet Light Trap, Review of Cinema, Widescreen Issue*, 21, 26–34, 1985.

Useful Websites

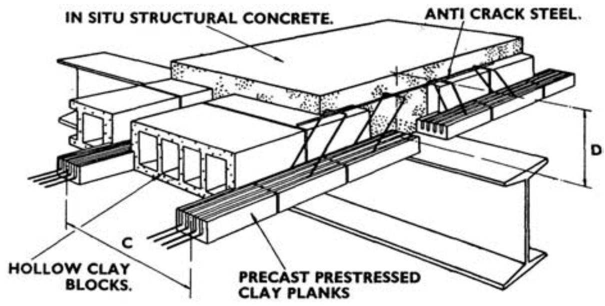
American Widescreen Museum website: http://www.widescreenmuseum.com/index.htm

Fire Engineering

The term "fire engineering" has gained growing acceptance in the construction industry only since the 1980s. However, the need for buildings that protected both the occupants and the structures themselves in case of fire has existed for 2000 years.

Major conflagrations (most famously, the Great Fire of London in 1666) often generated political pressures for legislation that required fireproof construction, though generally the object was saving buildings and their contents rather than lives. Until the late nineteenth century the word "fireproof" in the construction context was synonymous with ""incombustible." Throughout the nineteenth century an ever-increasing number of fireproof construction systems were patented for use in warehouses, factories and other large buildings. The earliest such system, from the 1790s, used brick or stone vaults (jack-arches) supported on cast-iron or, later, wrought iron beams and columns; this was used right up to the end of the nineteenth century, and many such buildings survive today. During the late nineteenth century systems generally incorporated iron beams in conjunction with concrete or hollow clay tiles.

However, the contents of factories and warehouses were often still highly flammable. Since much of the lighting in such buildings involved open flames, and cotton or wool fibers in the air formed a near explosive mixture, the conditions were nearly ideal for piloted ignition and spontaneous combustion. Severe fires and explosions were not uncommon and could lead to building collapse in two ways. An explosion could remove one or



FIRE RESISTANCE

Fire resistance	Additional finish required
½ hour	Nil
1 hour	Refer to makers
2 hours	⅜ inch gypsum plaster
4 hours	1 inch vermiculite plaster

Figure 3. Empirical approach to achieving fire protection. [*Source: British Constructional Steelwork Association.*]

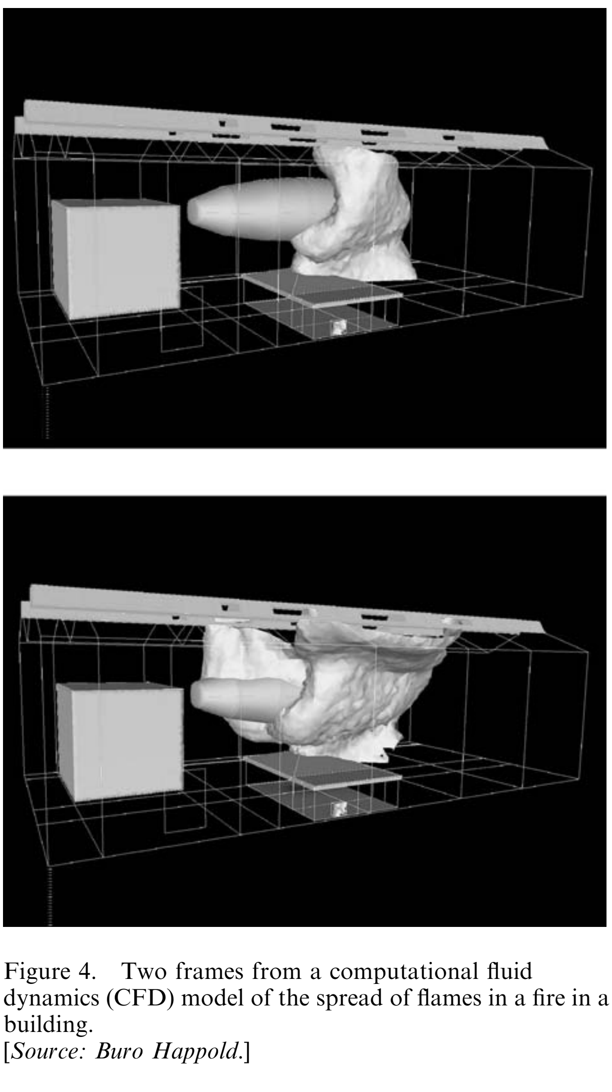


Figure 4. Two frames from a computational fluid dynamics (CFD) model of the spread of flames in a fire in a building.
[*Source: Buro Happold.*]

more vital structural members such as a column or load-bearing masonry wall. Alternatively, an intense fire could heat the exposed iron columns or beams to the point (about 550°C) where they lost virtually all their strength, and collapsed. Such was the technique of construction of these buildings that the removal of just one or two structural members could lead to a progressive and catastrophic collapse of a large part of the building, often causing many deaths. To such accidents were added a growing number of terrible fires in theaters in which the smoke from burning materials asphyxiated large numbers of the audience unable to escape through narrow corridors. By the end of the nineteenth century the whole idea of "fireproof" buildings had acquired a bad reputation and it was in this context that it was realized a new approach was needed to protecting buildings in fires – merely using incombustible construction materials was not sufficient.

Five main approaches were followed, almost simultaneously, during the early decades of the twentieth century. They all involved a logical, qualitative engineering approach to the problem and would be the necessary precursor to major quantitative developments later in the century. In contrast to the approach adopted in the previous century, these methods were tested experimentally in the growing number of fire research stations that were founded in many countries in the 1920s and 1930s:

1. The main cause of fire ignition was removed—electric lighting gradually replaced open flames.
2. Active means were installed to suppress or extinguish fires—portable fire extinguishers, an installed sprinkler system, or, in areas with electrical equipment, gas suppressants (usually carbon dioxide or Halon).
3. The idea of containment was introduced—ensuring that areas where a fire might start were isolated in a so-called "fire compartment," for instance through the use of self-closing doors or a fire safety curtain in theaters.
4. The exits from the occupied areas of buildings were also made larger, more accessible, and better signposted, for instance by providing fire escapes or isolated staircases.
5. Finally, steps were taken to retard the rate at which the iron or steel structural armature of the building heated up in a fire.

This latter idea was not new—structural timber had long been protected by plaster or even by iron

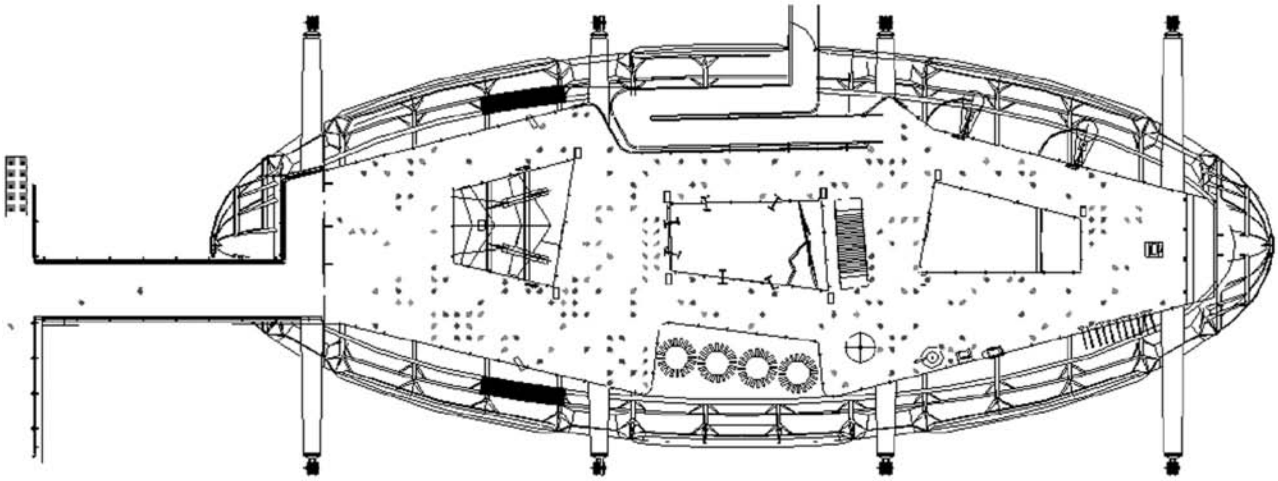


Figure 5. A frame from a dynamic model of people escaping from a building in a fire. Individual dots represent gender (male/female) and age (up to 30, 30–50 and > 50 years old). [*Source: Buro Happold.*]

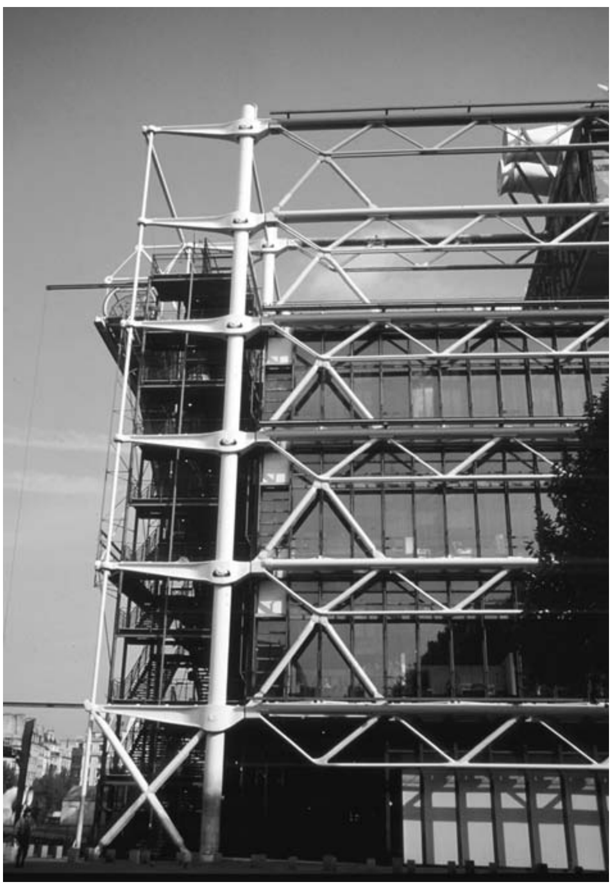


Figure 6. Pompidou Centre, Paris (1969–1974; engineered by Ove Arup and Partners). The exposed steelwork inside the building is fire-protected using intumescent paint. The main, vertical tubular columns are fire-protected by being filled with water, which would be pumped around to remove heat in a fire. [*Photo: Bill Addis.*]

sheets. It soon became a requirement to protect exposed steel and several methods were developed during the century—encasing columns in brickwork, covering or encasing beams in concrete, boxing the girders inside a case of insulating sheet made, for instance, from asbestos cement, and spraying a material containing a good insulator such as vermiculite to form an adhering layer up to 20 millimeters thick. The most recent development was intumescent paint which, when heated, foams and hardens to create a barrier of entrapped air bubbles which are highly insulating.

The aim of these different approaches was to achieve a certain fire resistance measured in units of time. A building element had to be made to survive the load of a standard fire test for a certain time, and the time depended upon the type of building (especially its height), and location of the element in the building. Typically the time might be 30 minutes for a two-story building or 60 minutes for a taller one. These times were deemed long enough for people to be evacuated and the fire itself to be suppressed. The whole issue of protecting the building structure thus usually became reduced to the relatively simple idea of providing sufficient "cover" to any load-bearing steel, whether a beam or the reinforcing bars in a reinforced concrete slab (Figure 3). However, as with many codified approaches to engineering design, this simplicity led to an unjustified feeling of comfort. Terrible fires still broke out and both collapses and deaths were still common. The matter was brought to a head in England by two terrible fires—the Summerland disaster in 1973 in

which 51 people, including many children, died in a fire in a leisure center, and a fire at Bradford football stadium in 1985. Both had been designed in accordance with current good practice.

These disasters provoked a fundamental rethink to achieving fire safety of building structures and, from the 1970s, an entirely new approach was developed. As already developed in other branches of engineering, it was based on mathematical models of the various phenomena involved. For the first time, attention was focused on the crucial parameter—the temperature of the metal in a building element subjected to a fire. The idea of a "fire load" was proposed in the 1920s by S.H. Ingberg and is typically measured in kilograms of wood; that is, the load on the building caused by burning so much timber. This, in many ways, is analogous to a gravity or wind load acting on a building. Knowing the thermal properties of the building materials, and with the newly available power of the computer, engineers were able to predict the thermal response of a structure in ways similar to predicting its structural response under gravity and wind loads. After much experimental testing to validate the mathematical models, it also soon became possible, for the first time, for engineers to model the behavior of fires themselves and thus understand and predict the behavior of buildings during fires. Computer modeling, in the form of computational fluid dynamics (CFD), can now also be used to study the temperatures of the building structure in a fire as well as the flow of hot gases in fires (Figure 4). The computer modeling of fires in buildings is now even able to model the movement or flow of people escaping from fires so that escape times can be more precisely predicted (Figure 5).

These various modeling techniques, together with a full risk analysis of a fire situation, are now collectively called "fire engineering" and represent what has been, perhaps, a quiet revolution in building design. Yet without it, we would not have the dramatic, exposed-steel structures that are now a relatively common sight. The Pompidou Center in Paris, conceived in the early 1970s, was one of the first such buildings (Figure 6). The ability to model the fire load and the structural response to this load allowed the design engineers to adopt the unusual idea of achieving fire resistance by filling the main columns with water which, in a fire, would be pumped around to remove heat from the steel to prevent it heating up too quickly. More common nowadays are the many buildings in which exposed steel can be used in a rather understated way, and the fire engineering approach to design can mean that the need for applied fire protection can be avoided altogether (Figure 7).

Figure 7. Bedfont Lakes near London (1993; engineered by Buro Happold). As a result of fire engineering design, none of the exposed steel needs fire protection.
[*Photo: Bill Addis.*]

See also **Concrete Shells; Constructed World**

W. A. Addis

Further Reading

Addis, W. Creativity and Innovation: The Structural Engineer's Contribution to Design. Architectural Press, Oxford, 2001.

Buchanan, A.H., Ed. *Fire Engineering Design Guide*. Centre for Advanced Engineering, University of Canterbury, New Zealand, 1994.

Drysdale, D. *Fire Dynamics*. Wiley, New York, 1998.

Hamilton S.B. *A Short History of the Structural Fire Protection of Buildings*. National Building Studies Special Report No. 27. Her Majesty's Stationery Office, London, 1958.

Ingberg, S.H. Tests of the severity of building fires. *National Fire Protection Association Quarterly, Boston*, 22, 1, 43–61, 1928.

Malhotra, H.L. *Fire Safety in Buildings*. Building Research Establishment, Garston, 1986.

Read, R.E.H. *A Short History of the Fire Research Station, Borehamwood*. Building Research Establishment, Garston, 1994.

Stollard, P. and Johnson, L. *Design Against Fire: An Introduction to Fire Safety Engineering Design*. Spon Press, London, 1994.

Fire Safety in Tall Buildings. Council on Tall Buildings and Urban Habitat, Lehigh University.

SFPE Handbook of Fire Protection Engineering, 2nd edn. SFPE & NFPA, New York, 1995.

Fish Farming

Controlled production, management, and harvesting of herbivorous and carnivorous fish has benefited from technology designed specifically for aquaculture. For centuries, humans have cultivated fish for dietary and economic benefits. Captive fish farming initially sustained local populations by supplementing wild fish harvests. Since the 1970s, aquaculture became a significant form of commercialized farming because wild fish populations declined due to overfishing and habitat deterioration. Growing human populations increased demand for reliable, consistent sources of fish suitable for consumption available throughout the year.

The United Nations Food and Agricultural Organization Food and Agriculture Organization (FAO) stated that fish farming productivity expanded more than 200 percent from 1985 to 1995. By the end of the twentieth century, fish farming provided approximately one third of fish eaten globally. Fish were a major protein source in developing countries where 85 percent of the world's fish farms exist. Farm-produced fish dominated agricultural production worldwide, outnumbering some types of livestock raised for meat. By 2000, the international aquaculture industry was estimated to value almost $50 billion.

In addition to contributing to markets, aquaculture created employment opportunities especially for women and unskilled laborers in Third World countries. Asia is the leading producer of farm-raised fish. China, where fish farming first occurred three centuries ago, grows 70 percent of farmed fish. In developing countries, fish farms offer people innovative means to enhance nutrition and to accrue profits. For example, South Pacific fish farmers drill holes in black-lipped oysters, *Pinctada margaritifera*, and submerge them on cords into water where the oysters form pearls around beads placed in their tissues. Because they offer alternative money sources, pearl farms reduce overfishing.

Technology appropriate for specific regions and species was crucial as fish farming industrialized. Aquaculture includes cultivation of freshwater and marine fish as well as mollusks, crustaceans, and aquatic plants. Fish farming relies on various engineering techniques to create functional artificial environments with suitable water quality to raise fish. Agricultural and civil engineers, chemists, and marine biologists collaborate, and the journal *Aquacultural Engineering* emphasizes this interdisciplinary approach.

Filtration, circulation, and oxygenation to provide adequate aeration and waste removal in tanks, man-made ponds, and pens are necessary to raise healthy fish. Some fish are cultivated in offshore floating pens. After they hatch, they are fed formula. Automated feeders and seal-scaring devices insure constant care. Fish farming protects fish from predators and minimizes exposure to diseases. Employees at the Dachigam National Park trout farm in Kashmir squeeze out eggs to hatch and raise fish in tanks prior to releasing them in the running waters of a contained stream.

Several fish farms have geothermally heated waters, which are warmed by power plant waste heat. Because the temperature of artificial fish habitats are regulated, fish can be raised indoors in facilities located in environments colder than their natural habitats. Indoor fish farming enables fish to be cultivated in urban areas. High yields can be produced in small spaces.

Initially, fish raised in captivity represented luxury species such as shrimp. In Europe and North and South America, salmon are raised in net pens. Other popular domesticated species include carp, cod, bass, and perch. Farm-raised fish are usually tastier and exhibit a greater consistency of size and quality than wild fish.

Because they are cold-killed in cool water, farm fish have a shelf life of at least two weeks. Fishing farms range in size and serve numerous purposes. Some backyard fish farms exist solely to feed owners' families, produce fish bait, generate extra revenues in local markets, or offer recreational fishing. Commercial fish farms raise fish for food, bait, aquarium pets, or for sport. Eggs, fry, and fingerlings are also sold to fisheries and to stock ponds. Oxygenated tanker trucks deliver fish to processors. Approximately 20 million trout are cultivated annually on U.K. fish farms. Catfish farming is widespread in the U.S.

Integrated fish farming (IFF) is a widely distributed technology. Farmers combine fish culture with other types of agriculture, such as raising rice or ducks either simultaneously or rotating cycles in the same water habitats as fish, especially when land and water resources are limited. In fields, fish live in furrows, pits, or ditches while rice grows on ridges. By utilizing farmland for multiple purposes, food and wastes are recycled and yields increase. Fish eat weeds and fertilize fields.

Biotechnology improves the quality of farm-raised fish by striving to produce hybrids and strains of fish with certain traits such as being disease-resistant and more resilient to cold climates. Researchers also apply technology to create genetically bigger fish that grow more quickly and need less oxygen to survive. Some transgenic fish produce more muscle from grain than their wild counterparts. The tilapia fish, indigenous to Africa, has been targeted for selective breeding experiments because it naturally thrives on wastes in poor habitats and requires minimal maintenance. The selectively bred strain reaches harvest maturity of more than 800 grams (1.75 pounds), has higher survival rates, and can reproduce when it is four months old, enabling as many as three harvests annually. The experimental channel catfish strain, NWAC-103, also attains maturity faster than other catfish. Conservation breeding is pursued to preserve endangered fish species. Fish breeding has demonstrated potential but lags behind genetic research and achievements applied to other livestock.

Computer technology has helped advance fish farming through the use of monitoring and security programs. Software such as FISHY tracks fish farm production. Electronic identification chips are used to tag farm fish. Satellite technology is used to determine climatic conditions such as weather patterns, atmospheric circulation, and oceanic tides that affect fish farms.

Agrichemicals serve as disinfectants, vaccines, and pharmaceuticals to combat bacterial infections, prevent disease, and kill parasites. Hormones are sometimes used to boost growth. Because carnivorous farm-raised fish consume large amounts of fish, researchers seek alternative feed sources to ease demands that have contributed to overfishing.

Fish farming technology can be problematic. If genetically engineered fish escape and mate with wild fish, the offspring might be unable to survive. Cultivated fish live in crowded tanks that sometimes cause suffocation, diseases, and immense amounts of waste and pollutants. Antibiotic use can sometimes result in resistant microorganisms. Coastal fish farms, especially those for shrimp, can be environmentally damaging if adjacent forests are razed.

See also **Breeding, Animal: Genetic Methods; Genetic Engineering, Applications**

ELIZABETH D. SCHAFER

Further Reading

Beveridge, MC.M., and McAndrew, B.J., Eds. *Tilapias: Biology and Exploitation*. Kluwer, Dordrecht, 2000.

Black, KD. and Pickering A.D., Eds. *Biology of Farmed Fish*. Sheffield Academic Press, Sheffield, 1998.

Ennion, SJ. and Goldspink, G., Eds. *Gene Expression and Manipulation in Aquatic Organisms*. Cambridge University Press, Cambridge, 1996.

Lahlou, B. and Vitiello, P., Eds. *Aquaculture: Fundamental and Applied Research*. American Geophysical Union, Washington D.C., 1993.

Mathias, J.A., Charles, A.T. and Bao-Tong, H., Eds. *Integrated Fish Farming*. CRC Press, Boca Raton, FL, 1998.

Pennell, W. and Barton, B.A., Eds. *Principles of Salmonid Culture*, Elsevier, New York, 1996.

Reinertsen, H., Dahle, L.A., Joergensen, L. and Tvinnereim, K., Eds. *Fish Farming Technology*. Balkema, Rotterdam, 1993.

Reinertsen, H. and Haaland, H., Eds. *Sustainable Fish Farming*. Balkema, Rotterdam, 1996.

Sedgwick, S.D. *Trout Farming Handbook*. Fishing News Books, Oxford, 1995.

Stickney, R.R. *Aquaculture in the United States: A Historical Survey*. Wiley, New York, 1996.

Treves-Brown, K.M. *Applied Fish Pharmacology*. Kluwer, Dordrecht, 2000.

Fission and Fusion Bombs

Fission weapons were developed first in the U.S., then in the Soviet Union, and later in Britain, France, China, India, and Pakistan. By the first decade of the twenty-first century, there were seven countries that announced that they had nuclear

weapons, and another three or four suspected of developing them.

The first atomic weapon tested was Fat Man, the plutonium weapon designed at Los Alamos and detonated at the Trinity test, Alamogordo, New Mexico, 16 July 1945. A weapon on the Fat Man design was dropped 8 August 1945 on the city of Nagasaki Japan, two days after the dropping of the uranium-fueled weapon, Little Boy, on Hiroshima. The Soviets developed their first nuclear weapon utilizing information gathered through espionage on the American program, much of it supplied by the German-born British physicist, Klaus Fuchs. The Soviets tested a duplicate of Fat Man in a test dubbed "Joe 1" by the American press in August 1949. The British cabinet decided to build nuclear weapons in January 1947, and the first British test was held in Australia in 1952. In the British program of testing weapons above ground, altogether there were 21 devices detonated between 1952 and 1958, 12 in or near Australia and 9 at Christmas Island. In addition to announced tests, the British tested many weapons components at Maralinga, the Australian test range.

Several designs of hydrogen bombs or thermonuclear bombs that relied on nuclear fusion rather than on fission for the majority of their energy release were considered in the period 1946–1955 by both the U.S. and the Soviet Union. In some designs, tritium, an isotope of hydrogen, was used to create a fusion effect that would more thoroughly cause fission in the plutonium core of the weapon, an effect known as "boosting."

In the late 1940s, American weapons designers thought through and abandoned the Alarm Clock design in which a fission weapon would ignite a deuterium fusion fuel. The device could be known as a fission–fusion–fission weapon, with layers of fissionable material, including some uranium-238 that would fission in the intense neutron environment. The upper limit of such a weapon would be several hundred kilotons. That design was not pursued or built.

The U.S. pursued the design of a cooled device using liquid deuterium, in the Ivy-Mike test of October 26, 1952. Later, the Teller–Ulam design worked out by Edward Teller and Stansilaw Ulam was far smaller and could be transported by aircraft. Teller and some others called the later design the "Classical Super."

The Soviets pursued a weapon similar to the U.S. Alarm Clock design, called the "Sloyka" or "Layer Cake" design, first conceived by Yakov Zel'dovich as the "First Idea" and improved by Andrei Sakharov. Sakharov's "Second Idea" involved surrounding the deuterium with a uranium-238 shell, in the layers or "sloyka," as Sakharov called it, to increase the neutron flux during the detonation. In a Russian pun, colleagues referred to the Layer Cake design as having been "sugarized," as the name Sakharov means "of sugar."

The Soviet scientists tested the Layer Cake with the Joe 4 test of 12 August 1953 that yielded 400 kilotons, with 15 to 20 percent of its power derived from fusion. The U.S. already had larger boosted weapons, and American analysis of the Joe 4 test concluded that the device in the Joe 4 test was simply a variation on a boosted fission weapon. At the Russian weapons museum at Arzamas, the Joe 4 weapon is labeled the world's first hydrogen bomb, while U.S. analysts have continued to regard it as a boosted fission weapon. Thus the question of whether a boosted weapon can be regarded as a fusion weapon is crucial to the issue of priority of invention.

The U.S. tested a new design that could be delivered as a weapon, with the Teller–Ulam design in the Castle/Bravo test held in 1954. That device yielded an estimated 15 megatons. The "Third Idea" developed by Igor Tamm and Sakharov, like the Teller-Ulam device, could go much higher in yield than the limits imposed by the Layer Cake idea, and the later Soviet tests of hydrogen bombs followed the Third Idea, with true, two-stage detonations. The first Soviet test of the Third Idea weapon came 22 November 1955 with a yield estimated at 1.6 megatons. Later models of this weapon had a yield of 3 megatons. The accuracy of some of this information cannot be confirmed from official sources, as the material is derived from generally scholarly but unauthorized open literature.

Some observers believed that Joe 4 represented a thermonuclear weapon, just as the Soviets asserted. If that had been true, it would appear that the Soviets developed a deliverable nuclear weapon about seven months before the U.S. did so with the 1954 Castle/Bravo test of the huge low-temperature fusion device, too large to carry as a weapon aboard an aircraft. Careful evaluation of the Joe 4 test revealed that it should be regarded as a boosted fission weapon, somewhat along the lines of that detonated in the Greenhouse/Item test by the United States in 1952. By such logic, American scientists could claim that the Classical Super test of Ivy/Mike in 1952 was the first true thermonuclear detonation, and that the Castle/Bravo Teller–Ulam weapon tested in 1954 was the world's

first deliverable thermonuclear weapon. That remains the American view of invention priority.

See also **Nuclear Reactors, Weapons Material; Warfare; Warfare, High Explosive Shells and Bombs**

RODNEY CARLISLE

Further Reading

Arnold, L. *A Very Special Relationship*. Her Majesty's Stationery Office, London, 1987.

Cochrane, T.B., *et al. Making the Russian Bomb:From Stalin to Yeltsin*. Westview Press, Boulder, CO, 1995.

Hewlett, R.G. and Anderson, O. The New World, 1939–1946, vol. 1: History of the United States Atomic Energy Commission. Atomic Energy Commission, Washington D.C., 1972.

Hewlett, R.G., and Duncan, F. *Atomic Shield, 1947–1952, vol. 2: History of the United States Atomic Energy Commission*. Atomic Energy Commission, Washington D.C., 1972.

Hewlett, R.G., and Holl, J.M. *Atoms for Peace and War, 1953–1961*. University of California Press, Berkeley, 1989.

Holloway, D. *Stalin and the Bomb*. Yale University Press, New Haven, 1994.

Rhodes, R. *The Making of the Atomic Bomb*. Simon & Schuster, New York, 1986.

Rhodes, R. *Dark Sun*. Simon & Schuster, New York, 1995.

Fly-by-Wire Systems

Most civilian and military aircraft are today controlled by "fly-by-wire" systems, in which controls are operated electrically, permitting greater maneuverability and safety than in conventional mechanically based control systems. Conventional flight control works as the pilot transmits directions through the control devices and the control surfaces are actuated according to the pilot's input. In larger airplanes, there is also a hydraulic system, like the power steering on an automobile, to actuate the control surfaces.

A fly-by-wire system breaks the mechanical lineages—a computer is added. Pilot input through the stick and rudder pedals are converted from a deflection to a voltage proportional to the deflection, and then the voltage is passed on to the computer, where it is further converted to binary numbers. Air data sensors are also connected to the computer. The computer then compares the pilot's desires to the actual state of the airplane and sends signals to reconcile them. These signals are converted back into voltages and used by electric servos at the hydraulics to move the control surfaces (see Figure 8).

Conventional systems depend on the pilot not doing something that is a bad idea, such as lifting the nose too much and causing a stall. In airplanes that are fly-by-wire, even if the pilot makes a mistake such as pulling up the nose too far or too fast, the computer can be programmed to prevent the airplane from responding in a dangerous way. This decoupling of the pilot from the airplane has to be lessened for fighter pilots, as they require increased maneuverability, not increased safety. Since many fly-by-wire airplanes are unstable in the pitch axis, it is simple to increase maneuverability using fly-by-wire.

During the first few decades of flight, airplanes were unstable. Originally there was a dispute about whether airplanes would be flown by "chauffeurs" or "pilots" The former would merely adjust the pointing of stable airframes, while the latter would have to successfully fight the instability of designs to keep the airplane in the air. At first, inventors who were would-be aviators adopted the chauffer model. When the Wright brothers realized that it was possible to pilot an unstable airplane, they flew. Their devices, plus most World War I fighters, were unstable and required great concentration to

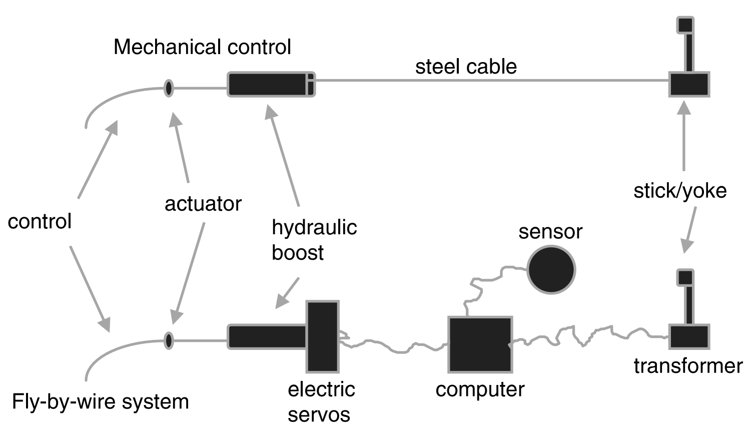


Figure 8. Fly-by-wire: how it works.

fly since the pilot's brain was the only "computer" and the seat of the pilot's pants the most important sensor on board. This was acceptable when fuel tanks were small and ranges limited. When airplanes were built to traverse long distances, they were designed to be stable in order to reduce the workload on their pilots. After decades of stable designs like these, once again the advantages of instability are available.

The Germans invented a sort of accidental fly-by-wire system to control part of their Mistel flying bomb during World War II. They filled worn-out Junkers-88 bombers with explosives, and a Messerschmidt-109 or Focke Wulf-190 fighter was then mounted on top. The fighter's pilot would steer this composite aircraft, a large three-engine biplane, to a target, aim it, and then release it to fly to the target. This was usually a dam or bridge with considerable concrete support. Junkers-88s were not commonly built with autopilots, but one was available. By placing potentiometers at the stick and rudder of the fighter, movements could be converted to electrical signals that were fed to the bomber's autopilot. Thus one man could fly both. The difference between the Mistel and a modern fly-by-wire aircraft is the presence of a computer.

Computers can be of two types: analog and digital. An analog computer operates on continuous functions, like voltage on a wire. A digital computer operates on discrete functions, so a voltage has to be converted to bits (ones and zeros) to be processed. In some ways, analog computers are better for airplane use, especially in that they accept continuous control inputs and do not need conversions. Digital computers are programmable with software and thus are more flexible. Analog computers have to be rebuilt to handle new flying characteristics; digital computers need only change programming. The first computers used in airplanes were analog. By the mid-1950s, most aeronautical engineers were convinced of the benefits of fly-by-wire. Only its practical implementation needed to be proved.

At about this time, Avro Canada was building a large, supersonic interceptor to help defend the country from Soviet nuclear attack by bombers. In American interceptors like the F-102 and the F-106, control passed to a ground-based site shortly after take-off. The ground controller would make the intercept, fire the weapons, and later pass control back to the pilot to complete the return to base and the landing. It would be far easier to make an interface with this type of system if the airplane's control system were electronic. Perhaps for this reason, or simply to achieve more maneuverability in their unwieldy airframe, the Canadians made the CF-105 Arrow fly by wire. The five prototypes of the Arrow made 66 flights in 1958–1959, so even though the fly-by-wire system was not perfected, this airplane featured fly-by-wire as it has come to be used.

In the 1960s in the U.S., the control division at Wright-Patterson Air Force Base sponsored several low-cost research projects aimed at making fly-by-wire possible. About the middle of the decade, it salvaged the tail section from a crashed B-47 and experimented with fly-by-wire in the pitch axis on the ground. A flying B-47 lofted this same pitch controller for some test flights. Short on money, Wright-Patterson tried fly-by-wire as a survivability improvement. A study of fighter-bombers that were damaged in Vietnam but returned safely to base found that the damage was not in a certain limited area of the airplanes where a major junction of control cables for the mechanical flight control system was located. Presumably those planes that were brought down were damaged in that area. By emphasizing survival ability, engineers obtained funding and an aircraft similar to the ones studied in Vietnam. This F-4 was converted to an analog fly-by-wire system but retained the mechanical system for the first 27 flights as a backup. It first flew on 29 April 1972 under control of the mechanical system and then switched to fly-by-wire in flight.

Meantime, encouraged by positive results with computer control of the Gemini and Apollo spacecraft, the National Aeronautics and Space Administration (NASA) of the U.S., took a surplus Apollo computer and pioneered digital fly-by-wire by installing the computer on an F-8. This airplane first flew under fly-by-wire control about a month after the U.S. Air Force's F-4.

Even though Boeing had a fly-by-wire system on its YC-14 cargo plane prototype, its competitor, Airbus, built the first commercial aircraft with fly-by-wire, the A320. The entire family of derivatives: the A318, A319, A330, and A340, have similar systems. Therefore, a technology first pioneered in the 1970s was in common use less than 20 years later, a highly rapid pace of change.

Aside from reliability, fly-by-wire offers improvements in safety. At least three redundant systems or two dissimilar systems are now part of every fly-by-wire flight control system. For example, even though NASA's first digital computer-controlled airplane used a single-string (one computer) system, even it had three of the analog computers like those in its contemporary plane, the U.S. Air Force's modified F-4, as a backup. The second system

installed on the NASA F-8 was triplex (three computers). This has been expanded to the point where the Boeing 777 has three strings each made up of three different computers, nine in all.

The Europeans have adopted a different method of redundancy. In Airbus fly-by-wire systems, there are two dissimilar strings. One controls the elevator and spoilers. The other controls the elevator and ailerons. The rudder is still mechanically controlled, as it is much less important to a commercial airplane. One of the systems has two computers; one has three. The machines contain software written by separate teams with dissimilar programming languages. The teams are even separated geographically. The basic idea is that two separate teams are unlikely to make the same errors, although this has not turned out to be the case. However, the dissimilar systems and redundant computers provide sufficient safety. Either method is a significant increase in overall safety.

See also **Aircraft Instrumentation; Warplanes, Fighters and Fighter Bombers**

JAMES E. TOMAYKO

Further Reading

Miller, F.L. and Emfinger, J.E. *Fly-by-Wire Techniques*. AFFDL-TR-67-53, U.S Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio.

Morris, J. *Background Fly-by-Wire Historical Information*. U.S. Air Force Flight Dynamics Laboratory Memo, Wright-Patterson AFB, Ohio.

Tomayko, J. *History of NASA's Digital Fly-by-Wire Project*. SP-4224, NASA Washington D.C., 2000.

Tomayko, J. *Code Name Mistletoe*. American Heritage of Invention and Technology, , 2000, pp. 26–33.

Tomayko, J. Blind faith: The United States Air Force and the development of fly-by-wire technology, in *Technology and the Air Force*. Government Printing Office, Washington D.C., 1997, pp. 163–185.

Tomayko, J. The Fly-By-Wire Revolution in Flight Control. *American Heritage of Invention and Technology*, 1992.

Foods Additives and Substitutes

Advances in food and agricultural technology have improved food safety and availability. Food technology includes techniques to preserve food and develop new products. Substances to preserve and enhance the appeal of foods are called food additives, and colorings fit into this category of additives that are intentionally included in a processed food. All coloring agents must be proven to be safe and their use in terms of permitted quantity, type of food that can have enhanced coloring, and final level is carefully controlled.

Fat substitutes on the other hand are technically known as replacers in that they replace the saturated and/or unsaturated fats that would normally be found in processed food as an ingredient or that would be added in formulation of a processed food. Usually the purpose is to improve the perceived health benefit of the particular food substance. Technically speaking, substitutes are not additives but their efficacy and safety must be demonstrated.

The incorporation of a food coloring or fat substitute does not confer the food with the status synthetic food. That description would normally be restricted to a food that was made purely from chemically synthesized components. The regulation of what is permitted in food is complex and is governed in the U.S. by the Food and Drug Administration. In the U.K. the role of regulating food and enforcing the regulations is now the responsibility of the Food Standards Agency, which works with other government departments such as the Department of Health to ensure that food is safe and labeled appropriately. In the U.S. the law defines "Standards of Identity" covering exactly what ingredients can be used and where. At one time this exempted labeling, but now all ingredients must be labeled.

All synthetic colors must be demonstrated to be safe before addition to food. Safety tests are carried out in animal species to determine the highest level of a substance that has no deleterious effect. In the U.S. the amount that is permitted is 1 percent of this level. Other measures of safety include acceptable daily intake (ADI) level, tolerable limits for environmental and chemical contamination in food, and maximum residue levels (MRL). Since food colors are dispensable, they are the most heavily criticized group of food additive. They do not improve food safety in any way, but simply change or enhance its appearance. Therefore it is possible to require that their use entails no extra risk to the consumer. With other additives such as preservatives, for example, there may be a necessary compromise between the risk of using them and the risk of not using them.

Color additives are also used to help identify flavors, for example a lemon-colored sweet should taste lemony, black jelly beans should taste like licorice, the soft center of a chocolate should taste of oranges if it is orange but of raspberries or strawberries if it is pink. The use of a new synthetic coloring agent requires an application to be submitted that describes the chemical structure, processing route, range of intended use, and final concentrations. Prior to the discovery of synthetic

dyes by Perkins (1856), only natural dyes were added to foods. Early synthetic dyes were metal based and included mercury and copper salts and arsenic poisons. Sometimes used to mask poor quality and even spoilage in food, many of these dyes were toxic. During the early part of the twentieth century a large number of cheaper dyes were synthesized, and many found their way into food products. Often these were complex chemicals and their physiological effect was not understood. These aniline or petroleum-based derivatives were covered by the U.S. Pure Food and Drug laws of 1906, which represented a landmark for food health and safety.

Once countries began to legislate what could be added to food, the number of natural and synthetic colors dropped markedly. By the late twentieth century there were major national differences in what is permitted. In Norway, the use of any synthetic coloring for food use has been banned since 1976. One source of great confusion is that natural dyes can be synthesized (the so-called nature-identical dyes). The coding for synthetic colors also underwent change; the E numbers for the European Union (EU) Code was being replaced by the Codex Alimentarius Commission system of coding, the so-called International Numbering System (INS). This largely follows the E numbers without the letter. In this system, tartrazine (C1 19140) is referred to as E102 in the U.K. and has an INS number of 102, but in the U.S. it is generally referred to as FD and C Yellow No. 5. It is a synthetic azo dye used in confectionery, fruit juices, canned fruits, brown sauces, pickles, pie fillings, and canned peas. It has an acceptable daily intake (ADI value) of 0 to 75 milligrams per kilogram of body weight. Tartrazine has been implicated in causing wakefulness in children and causing rhinitis, itching, and other allergenic symptoms. Individuals who wish to avoid tartrazine cannot simply avoid yellow-colored foods since tartrazine is also used to confer turquoise, green, and maroon colors. However U.S. and EU legislation requires that its use be listed on all food labels so that consumers can avoid it if they wish.

Amaranth E132 is a red food coloring made from synthetic coal-tar and azo dye. It was used in pudding mix, jelly crystals, and fruit fillings, but its use was banned in 1976 in the U.S. when a possible connection to malignant tumors was identified. However it is used in some countries with an average acceptable daily intake of 0 to 0.5 milligrams per kilogram of body weight as INS 123 (E123).

Approved food colorings are reviewed regularly and the ADI levels are amended accordingly. Reactions are rare (1 in 10,000 people are sensitive to tartrazine), but there is an Adverse Response Monitoring System (ARMS) in the U.S. and in most European countries.

Fat substitutes are compounds added in the formulation of complex foods that replace all or part of the fat and hence reduce the energy value and/or reduce the proportion of saturated fats. They also aid the digestion of foods. They must, by definition, function in the same way as the fat would in the body, for example act as binding agents or as humectants (giving food a moist texture). There are many fat substitutes available to the commercial food manufacturer. They are not required to be regulated as "novel foods" since the EU Novel Food Regulation (258/97) does not cover food processing aids. Low-fat spreads and cholesterol-lowering ingredients (containing phytosterols and phytostanols) are strictly speaking fat substitutes. Many of the low-fat spreads are whipped and have a high water content; hence they result in reduction in fat intake if they are replacing unmodified fat.

In addition to modifying food structure, providing more acceptable texture, imparting distinctive flavor, and adding to energy value, fats are often a source of fat-soluble vitamins (e.g., vitamins A and D). If an individual has an elevated cholesterol level, the use of a phytosterol or phytostanol spread to replace butter or margarine on bread will gradually reduce cholesterol levels. However, these products can cause gastrointestinal imbalance and are not recommended for babies, children, or those with conditions that might make them susceptible to vitamin A or D deficiency. One problem for consumers is that while it is clear what is in a package of these spreads, it is not always apparent which fats are incorporated in other foods such as yogurt, prepared baked goods, or restaurant meals. Hence it would be possible to unknowingly consume a daily intake above the recommended maximum. These issues remain of concern, and at the turn of the twenty-first century, the U.K. Food Standards Agency is looking at ways to monitor consumption of cholesterol-lowering spreads and prevent over consumption of specialized fats in diary goods by susceptible individuals.

See also **Food, Processed and Fast; Synthetic Foods, Mycoprotein and Hydrogenated Fats**

JANET BAINBRIDGE

Further Reading

Hochheiser, S. The evolution of U.S. food color standards, 1913–1919. *Agricult. Hist.*, 55, 4, 385–391, 1981.

Junod, SW. The chemogastric revolution and the regulation of food chemicals, in *Chemical Sciences in the Modern World*, Mauskopf, S., Ed. University of Pennsylvania Press, Philadelphia, 1993.

Useful Websites

U.S. Food and Drug Administration: Food Color Facts: http://vm.cfsam.fda.gov/-Ird/colorfac.html
FSA Website Homepage: www.food.gov.uk
FSA archive: www.archive.food.gov.uk

Food Preparation and Cooking

Twentieth century technological developments for preparing and cooking food consisted of both objects and techniques. Food engineers' primary objectives were to make kitchens more convenient and to reduce time and labor needed to produce meals. A variety of electric appliances were invented or their designs improved to supplement hand tools such as peelers, egg beaters, and grinders. By the close of the twentieth century, technological advancements transformed kitchens, the nucleus of many homes, into sophisticated centers of microchip-controlled devices. Cooking underwent a transition from being performed mainly for subsistence to often being an enjoyable hobby for many people.

Kitchen technology altered people's lives. The nineteenth-century Industrial Revolution had initiated the mechanization of homes. Cooks began to use precise measurements and temperatures to cook. Many people eagerly added gadgets to their kitchens, ranging from warming plates and toasters to tabletop cookers. Some architects designed kitchens with built-in cabinets, shelves, and convenient outlets to encourage appliance use. Because they usually cooked, women were the most directly affected by mechanical kitchen innovations. Their domestic roles were redefined as cooking required less time and was often accommodated by such amenities as built-in sinks and dishwashers. Ironically, machines often resulted in women receiving more demands to cook for events and activities because people no longer considered cooking to be an overwhelming chore.

Domestic technology contributed to home economics and dietetics becoming professional fields in the twentieth century. Home economists studied how household machines influenced cooking and advised people how to incorporate kitchen appliances into their lives. Guides instructed people how to utilize appliances to cook foods for entertaining groups. During the two world wars and economic depressions, people adjusted cooking techniques to cope with food shortages and rationing.

Throughout the twentieth century, inventors created new appliances to ease cooking burdens. When the century began, many kitchens in the U.S. still had wood-burning stoves or fireplaces. As electricity became available, some people invested in electric ranges. Costs limited mass acceptance, but gradually range costs became affordable. The standardization of electrical outlets, plugs, and currents in the 1920s and development of safety standards aided adoption of electric appliances. Electric stoves enabled cooks to bake goods without having to wait for a fire to warm sufficiently. These stoves were also cleaner than cooking on hearths. By the late 1970s, microwaves had replaced or supplemented stoves in many homes, altering how people prepared and cooked meals.

Cooks utilized a variety of appliances to prepare food for cooking. In the early twentieth century, engineers used small motors then magnetrons to create powerful kitchen appliances. Mixers quickly and smoothly combined dry goods with eggs, margarine, and other ingredients instead of people manually stirring dough. Crock pots and cookers enabled cooks to combine ingredients to cook unsupervised for a specified time. Automatic bread machines mixed, kneaded, raised, and baked breads. Coffee and tea makers brewed beverages timed for breakfast drinking and kept them warm. Espresso machines steamed frothy beverages. Cordless kettles heated liquids wherever people wanted to prepare hot drinks or soups. Juicers extracted liquid from fruits.

Some appliances were available only in certain geographical regions or met specific cultural needs such as rice steamers in Asia. As people traveled and encountered new devices, those technologies were often introduced to other countries. The Internet enabled people to become aware of and buy brands and types of appliances they might not find in local stores and are only available in specific countries or regions. Manufacturers such as Samsung and Toshiba produced appliances in Asia, while companies including DeLonghi and Bourgeois outfitted European homes. Innovators from many nations envisioned, adapted, and improved cooking tools.

Technologists worldwide created appliances to meet specific needs and local demand. The German manufacturer Miele produced the first electric dishwasher in Europe in 1929. Maurice Bourgeois

invented the first European built-in oven, resulting in his company becoming the leader in the convection oven market. At Sunbeam, Ivar Jepson designed kitchen appliances between the World Wars. The Sunbeam Mixmaster patented in the late 1920s surpassed other mechanical mixers because it had two beaters with interlocking blades that could be detached from the machine. Previously, the popular mixer that L.H. Hamilton, Chester Beach, and Fred Osius patented in 1911 only had one beater. Attachments enabled the Mixmaster to perform other tasks, including peeling, grinding, shelling, and juicing. The Mixmaster also could polish and sharpen utensils and open cans.

Inventors devised various electric toasters designs during the twentieth century. Efforts to create a reusable, unbreakable heating element to produce radiant heat for toasting sliced bread stymied many people. Engineer Albert Marsh patented Nichrome, a nickel and chromium alloy, in 1905. His invention enabled toaster heating elements to be produced. By 1909, consumers could purchase electric toasters developed by General Electric. Ten years later, the first pop-up toaster was patented. By the 1980s, toasters were designed to accommodate bagels. Plastics were used in addition to metals for cases, and microchip controls monitored toasting options.

Modern appliances often had pre-twentieth century precedents. Denis Papin designed the first pressure cooker in 1679 France. Later engineers adapted his cooker to produce an airtight environment in which steam cooked food so that vitamins and minerals were retained. In 1939, the National Pressure Cooker Company first sold a cast iron saucepan pressure cooker called Presto. After World War II, pressure cookers were made from stainless steel. By the late 1950s, engineers designed electric cookware, which had removable heat controls so that the pans, griddles, skillets, and coffee makers could be immersed in water to clean.

From the 1970s, cooks also used appliances designed to produce small servings. Sandwich machines and small indoor grills quickly cooked meals for individuals. Electric deep fryers prepared single portions of onion rings, french fries, and other fried foods. In contrast, kitchen technology also offered healthier fare. Hot-air popcorn poppers did not use oil. By the 1980s, the electric SaladShooter sliced and shredded ingredients directly into bowls.

Cooking technology benefited from inventors' curiosity. In 1938 at a DuPont laboratory, Roy J. Plunkett discovered Teflon while investigating chemical reactions occurring in tetrafluoroethylene (TFE), a refrigerant gas. Gas molecules had bonded to form polytetrafluoroethylene (PTFE) resin, which had lubricating properties, a high melting point, and was inert to chemicals. The process to create this polymer was refined and patented. When cooking pots and pans are coated with this polymer, they have non-stick surface that makes foods such as eggs and batters easier to cook.

In 1946, Earl Tupper invented Tupperware, which transformed how people stored and prepared food. These plastic containers with airtight seals were light and unbreakable, inspiring food technologists worldwide to use plastic instead of glass and metal materials. Tupperware can be used for cooking in microwaves and was environmentally sounder than disposable plastic and aluminum foil wraps.

In the latter twentieth century, microprocessors and materials such as polymers were used to make appliances lighter and easier to use. Engineers strived to make appliances smaller, more versatile and stable, quieter, and requiring less energy to operate. Digital technology made cooking more convenient because appliances with timers and sensors could be programmed to perform certain functions at specific times.

Radio programs featured cooking programs that advised cooks. Television introduced people to such notable cooks as Nigella Lawson, Raymond Oliver, Catherine Langeais, and Julia Child. Through the medium of television and video, cooks could demonstrate preparation methods such as basting and stuffing and cooking techniques including sautéing and frying that cookbooks often insufficiently described for inexperienced cooks to follow adequately. Television personalities posted recipes on web sites. Restaurants and food-related industries used the Internet to inform consumers how to make favorite meals at home and use specific products.

See also **Food Preservation, Cooling and Freezing; Food, Processed and Fast; Microwave Ovens**

ELIZABETH D. SCHAFER

Further Reading

Anderson, J. *The American Century Cook Book: The Most Popular Recipes of the 20th Century*. C. Potter, New York, 1997.

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.

Davidson, A. *The Oxford Companion to Food*. Oxford University Press, Oxford, 1999.

Du Vall, N. *Domestic Technology: A Chronology of Developments*. G.K. Hall, Boston, 1988.
Kiple, KF. and Ornelas, K.C., Eds. *The Cambridge World History of Food*, 2 vols. Cambridge University Press, Cambridge, 2000.
Macdonald, A.L. *Feminine Ingenuity: Women and Invention in America*. Ballantine Books, New York, 1992.
MacKenzie, D. and Wajcman, J., Eds. *The Social Shaping of Technology: How the Refrigerator Got Its Hum*. Open University Press, Philadelphia, 1985.
McBride, T.M. *The Domestic Revolution: The Modernization of Household Service in England and France, 1820–1920*. Croom Helm, London, 1976.
Plante, E.M. *The American Kitchen, 1700 to the Present: From Hearth to Highrise*. Facts on File, New York, 1995.
Shapiro, L. P*erfection Salad: Women and Cooking at the Turn of the Century*. Farrar, Straus & Giroux, New York, 1986.
Symons, M. *A History of Cooks and Cooking*. University of Illinois Press, Urbana, 2000.
Trescott, M.M., Eds. *Dynamos and Virgins Revisited: Women and Technological Change in History*. Scarecrow, Metuchen, NJ, 1979.
Williams, C., Ed. *Williams-Sonoma Kitchen Companion: The A to Z Everyday Cooking, Equipment and Ingredients*. Time-Life Books, Alexandria, VA, 2000.
Wolke, R.L. *What Einstein Told His Cook: Kitchen Science Explained.* W.W. Norton, New York, 2002.
Yarwood, D. *Five Hundred Years of Technology in the Home*. B.T. Batsford, London, 1983.

Food Preservation: Cooling and Freezing

People have long recognized the benefits of cooling and freezing perishable foods to preserve them and prevent spoilage and deterioration. These cold storage techniques, which impede bacterial activity, are popular means to protect food and enhance food safety and hygiene. The food industry has benefited from chilled food technology advancements during the twentieth century based on earlier observations. For several centuries, humans realized that evaporating salt water removed heat from substances. As a result food was cooled by placing it in brine. Cold storage in ice- or snow-packed spaces such as cellars and ice houses foreshadowed the invention of refrigerators and freezers.

Before mechanical refrigeration became consistent, freezing was the preferred food preservation technique because ice inhibited microorganisms. Freezing technology advanced to preserve food more efficiently with several processes. Blast freezing uses high-velocity air to freeze food for several hours in a tunnel. Refrigerated plates press and freeze food for thirty to ninety minutes in plate freezing. Belt freezing quickly freezes food in five minutes with air forced through a mesh belt. Cryogenic freezing involves liquid nitrogen or Freon absorbing food heat during several seconds of immersion.

In rural areas and small towns in the U.S., meat from butchering done by area residents was commonly frozen and stored in food lockers until after World War II. Refrigerators and freezers improved the quality of rural residents' lives and preserved milk from dairies, eggs from poultry, and meat from livestock. Worldwide, countries passed sanitation laws requiring specific refrigeration or freezing standards for foods. Refrigerators were miniaturized to become portable. Propane gas refrigerators were used to preserve food in recreational vehicles. By the end of the twentieth century, sophisticated refrigerators had several cooling zones within one unit to meet varying food needs such as chilling wine and crisping lettuce. Computer technology controlled these refrigerators' processes.

In the 1920s, government naturalist Clarence Birdseye (1886–1956) developed frozen foods by flash-freezing them. Inspired by watching Arctic people preserve meat in frozen seawater, Birdseye decided that quick freezing was why food remained edible. Fish cells were not damaged by ice crystals because freshly caught fish were frozen so quickly. Using ice, brine, and fans, Birdseye froze food rapidly in waxy boxes placed in a high-pressure environment. His freezing methods enabled foods to retain their flavors and textures. The foods' nutritional qualities were not altered. He established Birdseye Seafoods Inc., in 1924. Birdseye later sold his patents and trademarks for $22 million to the Goldman–Sachs Trading Corporation and Postum Company (later, the General Foods Corporation).

Birdseye's invention allowed people to consume a healthy diet of fresh vegetables and fruits out of season and at their convenience. He carefully marketed his frozen fare for retail sales. At Springfield, Massachusetts, consumers bought the first Birds Eye Frosted Foods. Despite the economic depression, consumers readily accepted buying frozen foods, including meat, vegetables, fruit, and seafood. Birdseye developed refrigerated display cases for grocery stores in 1930 and began manufacturing the cases four years later. By 1944, he shipped his frozen products in refrigerated boxcars, and they became popular throughout the U.S. His efforts served as a transportation and distribution model for other food companies.

Birdseye's inventiveness and marketing of frozen food inspired other food entrepreneurs globally. More types of food were frozen, including soups and pizza. Some frozen foods were packaged in foil. Frozen food production totaled 1.1 billion kilo-

grams annually during the 1940s. Governmental agencies such as the U.S. Food and Drug Administration set frozen food standards. The National Association of Frozen Food Packers (NAFFP), later renamed the American Frozen Food Institute, was established.

By the 1950s, almost two thirds of American grocery stores had frozen food sections. Precooked and boil-in-the-bag frozen foods joined frozen raw foods in displays. In 1954, Gerry Thomas first sold Swanson TV dinners. These complete meals in a tray that were heated in ovens appealed to consumers. Approximately ten million TV dinners were sold within the first year of availability. Commercially popular, frozen foods injected billions of dollars into the global economy. Airlines and restaurants relied on frozen foods for customers' meals. The Cold War motivated the U.S. Federal Civil Defense Administration to include frozen food in an atomic bomb test called Operation Cue, which determined that radiation did not affect frozen foods.

Frozen foods gained new popularity in the late twentieth century when fast-food restaurants such as McDonald's and Burger King prepared frozen items fresh for each customer. Increased variety, smaller servings for one person, low-calorie foods, and microwave ovens encouraged people to rely on frozen foods as meal staples especially as more women join the workforce. Government agencies publicly declared that frozen foods were nutritious. Many stores expanded refrigerated sections to meet customers' demands. Cryogenic railcars were invented to distribute frozen foods. Luxuries such as ice cream and chilled drinks have become commonplace due to refrigeration and freezer technology.

See also **Food Preparation and Cooking; Food Preservation: Freeze-Drying, Irradiation, and Vacuum Packing; Food, Processed and Fast; Transport, Foodstuffs**

ELIZABETH D. SCHAFER

Further Reading

Cleland, A.C. Food Refrigeration Processes: *Analysis, Design, and Simulation*. Elsevier, New York, 1990.

Dellino, C.V.J., Ed. *Cold and Chilled Storage Technology*. Van Nostrand Reinhold, New York, 1990.

Gormley, T.R., Ed. *Chilled Foods: The State of the Art*. Elsevier, New York, 1990.

Gormley, T.R. and Zeuthen, P. Eds. *Chilled Foods: The Ongoing Debate*. Elsevier, New York, 1990.

Mallett, C.P., Ed. *Frozen Food Technology*. Blackie Academic & Professional, London, 1993.

Stringer, M., and Dennis, C., Eds. *Chilled Foods: A Comprehensive Guide*, 2nd edn. CRC Press, Boca Raton, FL, 2000.

Wiley, R.C., Eds. *Minimally Processed Refrigerated Fruits and Vegetables*. Chapman & Hall, New York, 1994.

Food Preservation: Freeze-Drying, Irradiation, and Vacuum Packing

Humans have used processes associated with freeze-drying for centuries by placing foods at cooler high altitudes with low atmospheric pressure where water content is naturally vaporized. Also called lyophilization, freeze-drying involves moisture being removed from objects through sublimation. Modern freeze-drying techniques dehydrate frozen foods in vacuum chambers, which apply low pressure and cause vaporization.

Freeze-drying reduces foods' weight for storage and transportation. Freeze-dried foods do not require refrigeration but do need dry storage spaces. Adding water to freeze-dried foods reconstitutes them. Unlike dehydration, freeze-drying does not remove components that give foods flavors. The nutritional qualities of foods are also retained. The process inhibits microorganisms and chemical reactions causing food to spoil because the water pathogens and enzymes need to thrive is absent.

Engineers developed freeze-dryers for specific tasks that use vacuum pumps and chambers to pull moisture out of food. Commercially, freeze-dried foods are popular because they require less storage space than other packaged food. Freeze-drying extends the shelf life of products. Freeze-drying technology has been applied to consumer food products since the 1930s. Coffee, first freeze-dried in 1938, is the most familiar commercial freeze-dried food product. The Nestle Company was the first to freeze-dry coffee because Brazil requested assistance to deal with coffee bean surpluses. The successful freeze-drying of coffee resulted in the creation of powdered drinks such as Tang, which contrary to popular belief was invented by General Foods not the National Aeronautics and Space Administration (NASA). Many types of soups and noodles are freeze-dried. Researchers develop and patent new freeze-drying processes to improve and vary foods for consumers.

Freeze-drying was appropriated as the best method to preserve food for astronauts in space flight. During his pioneering February 1962 orbit, John Glenn was the first human to eat food in space. When he expressed his dissatisfaction, food engineers attempted to improve the taste of space

food while maintaining its nutritional qualities and minimizing its weight. As space technology advanced, more elaborate freeze-dried meals provided variety for longer duration space missions, which required compact, lightweight food cargo sufficient to feed crews at the international space station.

As early as 1963, radiation was used to control mold in wheat flour. The next year, white potatoes were irradiated to inhibit sprouting. By 1983, the Institute of Food Technologists released a study about the potential of radiation to sterilize and preserve food. Those experts declared that food processors might be reluctant to accept expensive irradiation technology unless they were confident that consumers would purchase irradiated goods. The Institute of Food Technologists warned that prices of irradiated food had to be affordably competitive with nonradiated foodstuffs and fulfill people's demands.

Irradiation is less successful than freeze-drying. Prior to irradiation, millions of people worldwide became ill annually due to contaminated foods with several thousand being hospitalized or dying due to food-borne pathogens. By exposing food to an electron beam, irradiation enhances food safety. Irradiated human and animal feed, especially grain, can be transported over distances and stored for a long duration without spoiling or posing contamination hazards. The radura is the international food packaging symbol for irradiation.

Small doses of radiation alters microbe DNA and kills approximately 99.9 percent of bacteria and parasites in meats. This exposure does not alter nutrients. Irradiation permits people to consume slightly cooked meats, including rare steaks and hamburgers. Most food irradiation uses cobalt-60 isotopes, but researchers developed alternative techniques such as gamma rays from cesium-137 and linear accelerators that transform electrons aimed at food into x-rays, which have greater penetration than electron beams. Those beams consist of high-energy electrons expelled from an electron gun and can only penetrate several centimeters compared to gamma and x-rays reaching depths of several feet. Irradiation sources are kept in water tanks that absorb the radiation until they are used to sterilize food in a thick concrete chamber.

Countries in North America, Europe, Asia, Africa, and the Middle East accepted irradiation. The World Health Organization endorsed irradiated food as safe for consumption. The U.S. Food and Drug Administration (FDA) approved irradiation of pork, fruit, vegetables, spices, and herbs in 1986 and poultry in 1990. Eight years later, the FDA approved the process of irradiating red meat to kill dangerous microorganisms and pests and slow meat spoilage. Despite federal approval, some states banned irradiation. During the 1990s, scientists improved irradiation methods to destroy toxins and bacteria, especially *Escherichia coli*, *Salmonella*, *Shigella*, and *Campylobacter*, in foods. The Institute of Food Technologists published a document that noted that the FDA's endorsement of irradiation legitimized that food preservation technology. Astronauts routinely eat irradiated food to prevent food-related sicknesses in space.

Despite irradiation's benefits, some consumers boycott purchasing or consuming irradiated foods that they consider dangerous and insist instead that facilities where agricultural goods are processed should be sanitized to reduce threats of contaminating toxins. Irradiation supporters assert that public opinion parallels how people initially reacted to milk pasteurization before accepting that process. Studies concerning how people and animals react to correctly irradiated food indicate that the foods are safe and not radioactive. Irradiation facilities are regulated by government licensing, and workers undergo rigorous training. Fatal accidents have occurred only when workers ignored rules regarding exposure to radioactive materials.

Irradiation has several significant limitations. Viruses are too small for irradiation dosages appropriate for safe food handling. Prions linked to bovine spongiform encephalopathy lack nucleic acid thus making irradiation ineffective.

Vacuum-packing food technologies involve a process that removes empty spaces around foods being packaged. Vacuum technology uses environments artificially modified to have atmospheric pressures that are lower than natural conditions. Vacuum packing extends the shelf life of food. The U.K. Advisory Committee on the Microbiological Safety of Foods warned that anaerobic pathogens such as *C. botulinum* can grow in vacuum-packed foods. Because vacuum packing often results in rubbery sliced cheese, some manufacturers use the modified atmosphere packaging (MAP) system, which utilizes gases to fill spaces so that cheese can mature to become tastier inside packaging.

See also **Food Preparation and Cooking; Food Preservation, Cooling and Freezing; Food, Processed and Fast; Transport, Foodstuffs**

ELIZABETH D. SCHAFER

Further Reading

Brennan, J.G. *Food Engineering Operations*, 3rd edn. Elsevier, London, 1990.

Dalgleish, J.M. *Freeze-Drying for the Food Industry*. Elsevier, London, 1990.

Jennings, T.A. *Lyophilization: Introduction and Basic Principles*. Interpharm Press, Englewood, CO, 1999.

Matz, S.A. *Technology of Food Product Development*. Pan-Tech International, McAllen, TX, 1994.

Rey, L. and May, J.C., Eds. *Freeze-Drying/Lyophilization of Pharmaceutical and Biological Products*. Marcel Dekker, New York, 1999.

Rotstein, E. Singh, R.P. and Valentas, K.J., Eds. *Handbook of Food Engineering Practice*. CRC Press, Boca Raton, FL, 1997.

Singh, R.P. and Heldman, D.R. *Introduction to Food Engineering*, 2nd edn. Academic Press, San Diego, 1993.

Wilkinson, V.M. and Gould, G.W. *Food Irradiation: A Reference Guide*. Butterworth-Heinemann, Oxford, 1996.

Food, Processed and Fast

Convenience, uniformity, predictability, affordability, and accessibility characterized twentieth-century processed and fast foods. Technology made mass-produced fast food possible by automating agricultural production and food processing. Globally, fast food provided a service for busy people who lacked time to buy groceries and cook their meals or could not afford the costs and time associated with eating traditional restaurant fare. As early as the nineteenth century, some cafeterias and restaurants, foreshadowing fast-food franchises, offered patrons self-service opportunities to select cooked and raw foods, such as meats and salads, from displays. Many modern cafeterias are affiliated with schools, businesses, and clubs to provide quick, cheap meals, often using processed foods and condiments, for students, employees, and members.

In the U.K., fish and chips shops introduced people to the possibilities of food being prepared and served quickly for restaurant dining or take-away. Sources credit the French with first preparing fried chips, also called fries, from potatoes, possibly in the seventeenth century. Simultaneously, the English bought fried fish at businesses that were so widespread that they were mentioned in nineteenth-century novels by writers including Charles Dickens. By the 1860s, English shopkeepers combined fish and chips for a nutritious and filling meal providing essential proteins and vitamins. Demand for fish and chips soared, with the approximately 8,500 modern shops, including the Harry Ramsden's chain, in the U.K. outnumbering McDonald's eight to one, and extending to serve patrons in countries worldwide.

Food-processing technology is designed primarily to standardize the food industry and produce food that is more flavorful and palatable for consumers and manageable and inexpensive for restaurant personnel. Food technologists develop better devices to improve the processing of food from slaughter or harvesting to presentation to diners. They are concerned with making food edible while extending the time period it can be consumed. Flavor, texture, and temperature retention of these foods when they are prepared for consumers are also sought in these processes. Microwave and radio frequency ovens process food quickly, consistently, and affordably. Microwaves are used to precook meats before they are frozen for later frying in fast-food restaurants. Nitrogen-based freezing systems have proven useful to process seafood, particularly shrimp. Mechanical and cryogenic systems also are used. The dehydrating and sterilizing of foods remove contaminants and make them easier to package. Heating and thawing eliminate bacteria to meet health codes. These processes are limited by associated expenses and occasional damage to foods. Processing techniques have been adapted to produce a greater variety of products from basic foods and have been automated to make production and packaging, such as mixing and bottling, efficient enough to meet consumer demand.

McDonald's is the most recognized fast-food brand name in the world. Approximately 28,000 McDonald's restaurants operated worldwide by the end of the twentieth century, with 2,000 opening annually. That chain originated in 1948 at San Bernadino, California, when brothers Richard and Maurice McDonald created the Speedee Service System. The thriving post-World War II economy encouraged a materialistic culture and population expansion. People embraced such technological developments as automobiles and enjoyed traveling within their communities and to distant destinations. The interstate highway system encouraged travel, and new forms of businesses catered to motorists. Drive-in restaurants provided convenient, quick meal sources.

The McDonalds innovated the assembly-line production of food. They limited their menus to several popular meals that could be consumed without utensils and were easy for children to handle. A hamburger, fries, and soft drink or milkshake composed the ubiquitous fast-food meal. Instead of hiring carhops to wait on customers, the McDonalds created a system in

which patrons served themselves. Employees did not need special skills to operate the McDonalds' food assembly line. As a result, the Speedee Service System reduced prices, establishing a family-friendly environment.

Multimixer milkshake machine salesman Ray Kroc bought rights to franchise the McDonalds' restaurant. He had traveled to the McDonalds' San Bernadino hamburger stand in 1954 because the brothers had bought eight milkshake machines. Because each machine had five spindles, 40 milkshakes could be produced at the same time. Curiously observing the stand, Kroc questioned customers about why they chose to patronize that restaurant. He then monitored kitchen activity and was especially intrigued by french fry preparations that created crispy yet soft fries. Kroc had an epiphany, deciding to establish a chain of identical hamburger stands where customers could expect food efficiently and consistently produced to taste the same regardless of their geographical location.

Technology was crucial for the spread of McDonald's. In addition to technical knowledge, tools, and materials to construct similar buildings which people would recognize as McDonald's anywhere, Kroc used technology to select suitable sites. He researched locations by flying above communities in airplanes or helicopters to determine where schools, shopping centers, and recreational places were located. Later, commercial satellites produced photographs for McDonald's representatives to identify where clusters of children and other possible customers were centralized. Computer software was utilized to analyze census and demographic information for optimal restaurant site selection.

Fast-food culture gradually altered eating habits in the latter twentieth century. Inspired by McDonalds' success, other major fast-food chains, including Burger King, Kentucky Fried Chicken, Wendy's, Domino's, Pizza Hut, Taco Bell, Dunkin' Donuts, and Carl's Jr. were created. Cake mixes, instant mashed potatoes, breakfast cereal, macaroni and cheese, and junk food such as potato chips, pretzels, canned cheese, and doughnuts also became popular processed and fast food available in grocery stores. Purchasing fast food became a routine activity for many people. The "drive-thru" enhanced fast food's convenience. In addition to restaurants, fast foods were served at airports, schools, gas stations, and large stores such as Wal-Mart.

McDonald's targeted children with marketing that promoted toys, often movie-related, in Happy Meals and featured children enjoying McDonald's playgrounds. McDonald's hired approximately one million workers annually. Most employees, many of them teenagers, were unskilled and paid minimum wages. Training was minimal, and the turnover rate was high. Fast-food corporations caused the demise of many independent food entrepreneurs. Conformity soon overshadowed regional cuisine.

Influencing agricultural distribution, McDonald's bought a large portion of the world's potatoes and meat. Slaughterhouses were industrialized to process hundreds of carcasses hourly on an assembly line system of conveyor belts and automated meat recovery (AMR) systems, which stripped all meat from bones. Because these machines often included bone and other contaminants in ground beef, European laws forbade their use. Meatpacking plants were cleaned with high-pressure hoses emitting a water and chlorine mixture. McDonald's implemented technological fast-food processes that altered how food was made. Only salad ingredients arrived at fast-food restaurants fresh. Everything else was reformulated and freeze-dried, dehydrated, frozen, or canned to prepare in fast-food kitchens designed by engineers. In laboratories, scientists manufactured chemicals to achieve desired flavors and smells in processed fast foods that were designed to please consumers.

Kroc hired people to assess water content of potato crops with hydrometers. He realized that potatoes needed to be stored in curing bins so that sugars converted into starches to prevent sugars from caramelizing during frying. Electrical engineer Louis Martino invented a potato computer for McDonald's to calculate how long fries should be cooked according to oil temperature. Fast-food chains used a system of blanching, drying, briefly frying, then freezing fries before they were deep fried for consumption. Sometimes fries were dipped in sugar or starch to achieve desired appearance and texture. At french fry factories, machines washed potatoes then blew off their skins. A water gun knife propelled potatoes 36 meters per second through a cutter that sliced them uniformly before they were frozen and shipped to restaurants.

Critics lambasted the high fat, sugar, salt, and calorie content of fast foods, which they linked to increased obesity, heart disease, and diabetes rates, especially in children. They demanded healthier options. Some menus were changed but were vulnerable to patrons' acceptance or rejection. Auburn University scientists replaced some fat in ground beef with carrageenan for moisture and flavor additives to create leaner beef that tasted like

normal ground beef. Marketed by McDonald's as the McLean Deluxe in the early 1990s, this lean hamburger failed to attract consumers because it was described as health food.

Hindus and vegetarians were angered when McDonald's disclosed that french fries were fried in beef tallow. Some people protested at the placement of fast-food restaurants in their neighborhoods. In 1990, McDonald's sued David Morris and Helen Steel, members of London Greenpeace, for libel because they distributed leaflets critical of the chain. The defendants lost the McLibel trial, but The Court of Appeals overturned some of the initial judgment. McDonald's stopped purchasing genetically engineered potatoes in an effort to prevent European consumer protests spreading to the U.S.

Fast-food globalization occurred as fast-food chains were built worldwide. McDonald's owned the most retail property internationally. In addition to archetypal fast-food meals, restaurants often accommodated local tastes such as Wendy's in Seoul, Korea, selling noodle dishes. Fast-food symbols were seen throughout the world. Ronald McDonald statues appeared in areas of the former East Germany, and the first McDonald's Golden Arch Hotel opened in Zurich, Switzerland, in 2001. A large plastic Colonel Sanders greeted passengers at Don Muang airport, Bangkok, Thailand. Even St. Kitts, a small Caribbean island, has a Kentucky Fried Chicken restaurant. Fast food was the target of anti-American protests in China, which internationally is second in the number of fast-food restaurants, and other countries.

See also **Food Preparation and Cooking; Food Preservation, Cooling and Freezing; Food Preservation: Freeze-Drying, Irradiation, and Vacuum Packing; Transport, Foodstuffs**

ELIZABETH D. SCHAFER

Further Reading

Alfino, M., Caputo, J.S. and Wynyard, R., Eds. *McDonaldization Revisited: Critical Essays on Consumer Culture*. Praeger, Westport, CT, 1998.

Barbosa-Cánovas, and Gould, G.W., Eds. *Innovations in Food Processing*. Technomic, Lancaster, PA, 2000.

Fishwick, M.W., Ed. *Ronald Revisited: The World of Ronald McDonald*. Bowling Green University Popular Press, Bowling Green, OH, 1983.

Hightower, J. *Eat Your Heart Out: Food Profiteering in America*. Crown Publishers, New York, 1975.

Jakle, J.A. and Sculle, K.A. *Fast Food: Roadside Restaurants in the Automobile Age*. Johns Hopkins University Press, Baltimore, MD, 1999.

Kincheloe, J.L. *The Sign of the Burger: McDonald's and the Culture of Power*. Temple University Press, Philadelphia, PA, 2002.

Kroc, R. and Anderson, R. *Grinding It Out: The Making of McDonald's*. H. Regnery, Chicago, 1977.

Lee, T.-C. and Ho, C.-T., Eds. *Bioactive Compounds in Foods: Effects of Processing and Storage*. American Chemical Society, Washington D.C., 2002.

Lusas, E.W. and Rooney, L.W. *Snack Foods Processing*. Technomic, Lancaster, PA, 2001.

McLamore, J.W. *The Burger King: Jim McLamore and the Building of an Empire*. McGraw-Hill, New York, 1998.

Moreira, R.G. *Automatic Control for Food Processing Systems*. Aspen, Gaithersburg, ME, 2001.

Ritzer, G., Ed. *McDonaldization: The Reade*r. Pine Forge Press, Thousand Oaks, CA, 2002.

Royle, T. and Towers, B. *Labour Relations in the Global Fast Food Industry*. Routledge, New York, 2002.

Rozin, E. *The Primal Cheeseburger*. Penguin, New York, 1994.

Schlosser, E. *Fast Food Nation: The Dark Side of the All-American Meal*. Houghton Mifflin, New York, 2001.

Smart, B., Ed. *Resisting McDonaldization*. Sage, London, 1999.

Vidal, J. *McLibel: Burger Culture on Trial*. New Press, New York, 1998.

Watson, J.L., Ed. *Golden Arches East: McDonald's in East Asia*. Stanford University Press, Stanford, CA, 1997.

Fossil Fuel Power Stations

Until the last third of the twentieth century, fossil fuels—coal, oil, natural gas—were the primary source of energy in the industrialized world. Large thermal power stations supplied from fossil fuel resources have capacities ranging up to 4000 to 5000 megawatts. Gas-fired combined-cycle power stations tend to be somewhat smaller, perhaps no larger than 1000 to 1500 megawatts in capacity.

Concerns in the 1970s over degradation of urban air quality due to particulate emissions and acid rain from sulfur dioxide emissions from fossil fuel power stations were joined from the 1990s by an awareness of the potential global warming effect of greenhouse gases such as carbon dioxide (CO_2), produced from the combustion of fossil fuels (see Electricity Generation and the Environment). However, despite a move towards carbon-free electricity generation, for example from nuclear power stations and wind and solar plants, fossil fuels remain the most significant source of electrical energy generation.

Basic Technology

Power station technology for converting fuel into useful electricity comprises two basic stages:

1. Combustion, either of fuel in a boiler to raise steam or gas or distillate in a gas turbine to produce direct mechanical power
2. Conversion of the energy in steam in a steam turbine to produce mechanical power

In both cases, the mechanical power is used to drive an electricity generator. The more efficient steam turbines developed by Charles Parsons and Charles Curtis at the end of the nineteenth century replaced the use of reciprocating (i.e. piston-driven) steam engines to drive the dynamo.

Efficiency in the fuel-to-electricity conversion process for large coal and oil-fired power stations is seldom greater than about 36 to 38 percent. Combined-cycle gas turbine stations on the other hand can have efficiencies around 60 percent.

Coal

Coal was the source of energy for the Industrial Revolution in the U.K. and for the rapid expansion of industry in the eighteenth and nineteenth centuries. Since the middle of the twentieth century, the introduction of other more convenient fuels than that of the labor-intensive coal industry has caused its decline throughout Western Europe. Coal as a fuel dominated the power industry in the first half of the twentieth century but in the U.K., its use for power generation has declined from 90 percent in 1950 to 15 percent in 2000. In the U.S., coal has become more important as other fossil fuel supplies decline and large coal reserves remain. In 1995, coal burning produced about 55 percent of the electricity generated in the U.S. The largest coal power station in Europe is situated at Drax in Yorkshire and has a capacity of 4000 megawatts.

In the 1920s, pulverized coal firing was developed, improving thermal efficiency at higher combustion temperatures. Powdered coal is suspended in air and blown directly into the burner. Cyclone furnaces, developed in the 1940s, used crushed coal and an injection of heated air, causing the coal-air mixture to swirl like a cyclone. This rotation produced higher heat densities and allowed the combustion of poorer grade of coal with less ash production and greater overall efficiency.

Electrostatic precipitators, filtration, or wet scrubbing (quenching the raw fuel gas with water) is used to reduce particulate emissions in flue gases. In the electrostatic precipitator, an ionizing field imparts an electric charge to the particles, allowing them to be collected on an oppositely charged surface.

Oil

From about 1945 onward, the rapid increase in the quantities of crude oil refined has given rise to large quantities of heavy fuel oil (HFO) being produced, which can be burned in large power stations to raise steam for turbine-generators. For many years, the cost of oil (on an equivalent heat basis) was lower than coal, and this gave rise to the construction of large oil-fired power stations, generally sited close to oil refineries on the coast. HFO can contain significant quantities of sulfur, which causes severe atmospheric pollution unless the sulfur oxide flue gases are subjected to chemical treatment (desulfurization). The price of HFO is prone to considerable fluctuation on the world market, and this can affect its competitive position as a power station fuel. Many power stations around the world, particularly medium-scale ones (up to 50 megawatt capacity), utilize distillate or "diesel" fuel to power large reciprocating engines which drive electricity generators. Such liquid fuel can be used also to power gas turbines to drive generators.

Natural Gas

Natural gas (methane) is generally associated with oil reserves and can be recovered for use as a domestic, industrial and power station fuel. While it can be burned in conventional power station boilers to raise steam for turbine-generators, it is more usually utilized and burned directly in gas turbines, which drive electrical generators (see Turbines, Gas). The efficiency of this method of power production can be raised significantly (to over 60 percent) by recovering the waste heat from the gas turbine exhaust and using this to raise steam to drive a separate turbine generator set. Most of the new power stations constructed in the developed world in the latter part of the twentieth century have been based on this so-called "combined-cycle" arrangement. Such power stations are generally cheaper, quicker to build, and require less land area than a conventional coal or oil-fired power station of equivalent capacity.

More efficient gas turbines have resulted in the energy-efficiency factor for gas increasing by 40 percent from the early 1980s to the late 1990s. New natural-gas-powered plants also have much reduced sulfur dioxide (SO_2) and nitrous oxides (NO_x) emissions compared to pulverized-coal-fired steam plant, even those with "scrubbers" that remove polluting gases and particulates.

Orimulsion and Petroleum Coke

Orimulsion is a bitumen-based fuel of which there are large reserves in North and South America and elsewhere. It is widely regarded by environmentalists as a "dirty" fuel, and thus it finds little

application as a power station fuel at the beginning of the twenty-first century. It has enormous potential when used with appropriate emission-control equipment. Petroleum coke is a byproduct of oil refining and can be gasified to produce a synthetic gas (Syngas) for power generation purposes.

Delivery and Storage of Fuel

Continuous supplies of fuel are essential for any electricity generating plant. Coal tends to be delivered by railway trains or by ship. Gas or oil would be delivered by pipeline, either from a country's "gas grid" or from a marine terminal or refinery. Large quantities of coal or oil fuel are usually stored on a power station site to cover for periods when deliveries might be interrupted. An alternative or "back-up" fuel such as distillate would be stored at a gas-fired power station.

Siting of Power Stations

The three critical factors that determine the location of a large central power station are source of fuel, supplies of cooling water, and proximity of load.

In the case of a coal-fired power station, in Britain the site may be on or very close to a coal field or coal mine. Sometimes, large coal power stations might be situated on an estuary and have coal delivered via ships. Rail links might also be used to deliver the coal, using large 1000-ton capacity freight trains to a specially constructed rail head at the power station.

Oil-fired power stations tend to be sited adjacent to an oil refinery, the fuel being conveyed directly from the refinery to the power station by pipeline. Sometimes a location on an estuary might be selected and the fuel oil delivered to offshore jetties by large oil tankers.

Gas-fired power stations are built so as to gain access to the gas transmission network in a country and may also be located close to the point of gas production; for example, on the North Sea coast in the U.K.

Where the power station utilizes steam turbines to drive the electrical generators, large supplies of cooling water are required to condense the steam back to water after it leaves the steam turbine. This condensate is then circulated back to the boilers where it is converted into steam once again. Where rivers or the sea are not available or, for environmental reasons cannot be used as sources of cooling water, cooling towers may be used which serve the same condensing purposes, albeit at lower overall efficiency.

In the early days of the electricity industry, power stations were located in the middle of towns and cities and the electricity consumers were served by cables radiating from the power station. As stations became larger, the lack of suitable large sites tended to result in their siting in areas where either or both of fuel and cooling water were obtainable. The electricity was transmitted to the load center by a "grid" of high-voltage power lines. It is generally more economical to convey the electrical energy by high-voltage power lines from the power station over long distances than to transport the fuel by road, rail or sea over the same distance.

Where gas-fired power stations are established in heavily built-up regions like the U.K., the existence of a large gas transmission network over most of the country means that the stations, because of their relatively compact size, can be sited close to the load centers and thus reduce the requirements for high-voltage power lines.

See also **Electrical Energy Generation and Supply, Large Scale; Electricity Generation and the Environment; Energy and Power**

IAN BURDON

Further Reading

Edgerton, D. *Science, Technology and the British Industrial 'Decline' ca. 1870–1970*. Cambridge University Press/ Economic History Society, Cambridge, 1996.

Elliot, D. *Energy, Society and Environment*. Routledge, London 1997.

Leggett, J. *The Carbon War: Global Warming at the End of the Oil Era*. Penguin, London, 1999.

National Research Council. *Coal: Energy for the Future*. National Academy Press, Washington D.C., 1995.

Reddish, A. and Rand, M. The environmental effects of present energy policies, in *Energy, Resources and Environment*, Blunden, J., and Reddish, A., Eds. Hodder & Stoughten (in association with The Open University, London, 1991.

Wollenberg, B.F. *Power Generation, Operation and Control*. Wiley/Interscience, New York, 1996.

Useful Websites

U.S. Department for Energy, Office of Fossil Energy: http://www.fe.doe.gov/

Fuel Cells

The principle of the fuel cell (FC) is similar to that of the electrical storage battery. However, whereas the battery has a fixed stock of chemical reactants and can "run down," the fuel cell is continuously

supplied (from a separate tank) with a stream of oxidizer and fuel from which it generates electricity. Electrolysis—in which passage of an electrical current through water decomposes it into its constituents, H_2 and O_2—was still novel in 1839 when a young Welsh lawyer–scientist, William Grove, demonstrated that it could be made to run in reverse. That is, if the H_2 and O_2 were not driven off but rather allowed to recombine in the presence of the electrodes, the result was water—and electrical current.

The mechanism of electrical generation was not fully understood until development of atomic theory in the twentieth century. Like a battery, an FC has two electrodes connected via an electrical load and physically separated by an electrolyte—a substance that will selectively pass either positive ions (acidic electrolyte) or negative ions (alkaline electrolyte). Oxidizer (e.g., O_2) enters at one electrode and fuel (e.g., H_2) at the other. At the cathode (the electrically positive electrode) atoms of fuel occasionally ionize, forming positively charged ions and free electrons. This is accelerated by a catalyst at the electrodes as well as suitable conditions of temperature and pressure. Similarly at the negative electrode (anode), oxidizer atoms spontaneously form negative ions. Then, depending on electrolyte, either positive or negative ions migrate through it to combine with ions of opposite polarity, while electrons (unable to pass through the electrolyte) flow through the electrical load from anode to cathode.

While the Second Law of Thermodynamics severely limits the efficiency of heat engines operating at practical temperatures, fuel cells can extract more power out of the same quantity of fuel compared to traditional combustion, since the hydrogen fuel is not converted to thermal energy, but used directly to produce mechanical energy. In principle, an FC consuming pure hydrogen and oxygen and producing liquid water can achieve an efficiency of 94.5 percent with an open-circuit potential of 1.185 volts. In practice a variety of losses cannot altogether be eliminated, but achievable efficiencies are considerably greater than those of most heat engines. Multiple cells connected in series supply higher voltages.

Practical challenges include electrolyte–electrode chemistry and physical properties, catalysts, fuel and oxidizer composition and purity, internal electrical losses, corrosion and other destructive chemical reactions, and a host of mechanical issues. Efficiency, silence, and lack of polluting emissions stimulated great interest, however. In searching for suitable systems, twentieth century researchers developed various approaches, most named for their electrolytes. Most of the developments in the field have come as a result of proprietary interests and the work of corporate teams of researchers.

Alkaline FCs (AFCs)

From the 1930s Francis Bacon of the U.K., followed later by U.S. researchers, turned from acidic electrolytes to more tractable potassium hydroxide (KOH). Bacon also pioneered use of porous electrodes through which gaseous reactants diffused. The first major application of FC technology was in the U.S. space program where AFCs provide power (and potable water) for the space shuttles. The need for extremely pure H_2 and O_2 made AFCs uneconomic for more mundane applications.

Molten carbonate FCs (MCFCs)

MCFCs grew out of SOFC research (see below). In the 1950s Dutch investigators G.H.J. Broers and J.A.A. Ketelaar turned to molten lithium-, sodium, and potassium carbonates as electrolytes, as did Bacon in the U.K. At typical 650°C operating temperatures, MCFCs produce waste heat in a form useful for industrial purposes or to power turbines for added electrical output. High temperatures relax the need for costly catalysts while their carbonate chemistry is tolerant of carbon monoxide (CO), which as an impurity is problematic for alkaline fuel cells. However chemical and mechanical problems have thus far impeded wide application.

Phosphoric Acid FCs (PAFCs)

Interest in phosphoric acid as an electrolyte emerged slowly until the mid-1960s, after which PAFCs rapidly became the first FCs to see significant commercialization. At around 200°C, PAFC waste heat may be used to reform (convert) hydrocarbons or coal to H_2 for fuel or power an auxiliary turbine. PAFCs are relatively tolerant of CO but sulfur must be separated. Units up to 250 kilowatts output are sold for fixed-site power applications and experimental units have shown promise in buses. Problems center on internal chemical reactions and corrosion.

Proton-Exchange Membrane FCs (PEMFCs)

In the early 1960s Thomas Grubb and Leonard Niedrach of General Electric in the U.S. developed

a polymer membrane which, moistened with water, served as an effective and stable electrolyte. Initial space application attempts in the 1960s revealed reliability problems (and led to adoption of AFCs) but further development has held out strong promise for ground and marine vehicle applications as well as small fixed generators. Operating temperatures of less than 100°C and high power relative to size and weight, suit PEMFCs especially. Platinum catalysts are necessary and the H_2 fuel must be essentially free of CO. By century's end PEMFCs had shown some promise of being adaptable to liquid methanol in place of H_2 fuel.

Solid-Oxide FCs (SOFCs)

Beginning in the 1930s experimenters first in Switzerland and Russia sought high-temperature (around 1000°C) solid ceramic electrolytes. At such temperatures, reactions proceed rapidly without costly catalysts and many fuel stocks can be reformed to produce H_2 within the SOFC, while the waste heat can be used for many purposes. But physical and chemical problems of high-temperature operation have been difficult and it remained uncertain at century's end how well the promise of SOFCs could be realized.

Metal FCs (MFCs)

To avoid problems of H_2 supply or conversion some FC developers turned, late in the century, to metal fuels, usually zinc or aluminum. Electrolytes include liquid KOH and proton-exchange membranes. Waste products are metal oxides.

Conclusion

Over the twentieth century FCs moved from laboratory curiosity to practical application in limited roles and quantities. It is very possible that the twenty-first century will see them assume a major or even dominant position as power sources in a broad array of applications. Obstacles are largely economic and the outcome will be influenced by success in development of competing systems as well as FCs themselves.

See also **Batteries; Electrochemistry**

William D. O'Neil

Further Reading

The future of fuel cells. *Scientific American*, 1999.
Kartha, S. and Grimes, P. Fuel cells: energy conversion for the next century. *Physics Today*, 1994.
Montavalli, J. *Forward Drive: The Race to Build "Clean" Cars for the Future*. Sierra Club Books, 2000.
Riezenman, M.J. Metal fuel cells. *IEEE Spectrum*, 38, 2001.
Srinivasan, S. *et al*. Fuel cells: reaching the era of clean and efficient power generation in the twenty-first century. *Ann. Rev. of Energ. Environ.*, 24, 1999.

Useful Websites

Smithsonian Institution Fuel Cells History Project, http://fuelcells.si.edu/index.htm

Fusion, *see* **Nuclear Reactors: Fusion**

G

Gender and Technology

As a field of scholarship, the study of gender and technology evolved rapidly in the latter half of the twentieth century. The field began with the study of women in technology, focusing on the leading women in engineering or female inventors, who were found primarily in Western societies. These studies were often in a biographical and historical mode, leading to an appreciation of the barriers women faced and overcame to join technical professions. However, scholars came to understand that this privileged the conventional or commonsense definitions of technology, which was equated with large-scale, complex, and public enterprises such as aerospace, railway, and civil engineering projects. Cowan (1983) and McGaw (1996) argue that this realization led to scholarship that broadened the definition of technology to include what had often been stereotyped as "women's work" and which included such items as domestic appliances, typewriters and other tools of clerical labor, and artifacts of personal care. In these kinds of studies, the technical complexity of women's activities, including invention, adaptation, and the use of artifacts, was highlighted and understood as being interconnected with larger social and economic systems.

Concern with looking at women in technology and looking at women's technologies was informed by what is known as the "second wave" of feminism in the 1970s and later. This period reflected the dramatic changes in the status of women, starting first with the experiences of women in the workforce during both World War I and II and continuing through the radicalization of feminism in the context of the 1960s civil rights era in the U.S. In these perspectives, issues of equal access and equal treatment in major social institutions were concerns of both scholarship and activism. This inclusive approach was also fueled by the launch of Sputnik in the USSR, which focused attention on U.S. competitiveness in space and science more generally, and spurred an interest in science and engineering for both young men and women. The historical studies were soon joined by sociological and anthropological inquiries such as those raised by Witz (1992) and Cockburn and Ormrod (1993) that asked questions about how it is that work is seen to be appropriate for people of one gender or the other and how technologies come to be associated with a particular gender. Sociologists such as McIlwee and Robinson (1992), Ranson and Reeves (1996) and Roos and Gatta (1999) have studied women in technological industries and workplaces and examined different patterns of compensation and promotion, both in terms of discrimination and success. This scholarship illustrated that while some gains had been made, promotion and compensation rates for women lagged those of men even when factoring in years of experience and child-rearing breaks in career paths.

Late twentieth century research also looked at cross-national comparisons of the professional engineering sector, as well as studies of how development projects differentially affect men and women in nonindustrialized countries. For example, while engineering is highly gender segregated in the U.S., other countries have more integrated engineering programs (the former Soviet Union and Italy) and computer science and telecommunications employment fields (as in Southeast Asia). International development projects often affect

women's livelihoods in nonindustrialized societies, sometimes improving their social position but often having a negative impact on work, property ownership, and family health.

This kind of complexity means that it is hard to generalize about exactly how technological change has influenced the roles of men and women. In some cases, technological change opened new avenues for work and economic mobility, as in the early days of clerical labor fostered by the typewriter or the early phases of the computer revolution. Frequently, however, the pattern of technological change followed existing lines of power in society and reinforced systems of stratification. For example, Ruth Schwartz Cowan argued that while "labor saving" household technologies such as washers and dryers or dishwashers save women from direct physical labor, women still spent nearly as much time as they did a century ago on housework, as well as handling the additional demands of working outside the home.

At the same time that scholars have made the definition of technology more inclusive and complex, there is increased understanding of the relationship between identity and work. Butler (1993) and Connell (1995) argue that how identities are defined by individuals and society, how they change over time, and how they vary with regard to race, class and ethnicity, have helped shape configurations of masculinity and femininity. The study of identity and difference has shown that gender is a way that cultures categorize the perceived differences between men and women. Thus the study of women and technology has been transformed into the study of gender and technology, the study of ideas about masculinity and femininity, and the inseparability of these concepts from cultural ideas about technology and work. This development of the understanding that gender is a constellation of values and ideas that any given culture associates with sexual differences is socially constructed and situationally produced and is not a natural category. Similarly, what is considered technology is a classification bounded by cultural ideas about gender, status, and skill. Croissant (2003) argues that in much the same ways that classical or "high" art is distinguished from crafts and hobbies, technology—particularly "high tech"—is a way of making distinctions of worth and value based on cultural stereotypes that overlap with gender systems. So, as scholars such as Oldenziel (1999) and Pacey (1983) have argued, ideas about technology in Western culture are strongly associated with masculinity, particularly a middle-class professional identity associated with engineering.

Several late twentieth century scholars such as McGaw (1996) and Croissant (2000), hoped that an understanding of how gender ideas help shape ideas about technology would influence people to look beyond stereotypical ideas about both technology and gender. This could result in opening employment opportunities for men and women and increase the diversity of engineering and related technical work. In addition, such an understanding could provide an important set of tools for critical thinking about the interconnections of our technological society.

JENNIFER CROISSANT

Further Reading

Ambrose, S.A. *Journeys of Women in Science and Engineering: No Universal Constants.* Temple University Press, Philadelphia, 1997.

Butler, J. *Bodies that Matter: On the Discursive Limits of Sex*. Routledge, New York, Routledge, 1993.

Cockburn, C. and Ormrod, S. *Gender and Technology in the Making*. Sage, London, 1993.

Connell, R.W. *Masculinities*. Polity Press, Cambridge, 1995.

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.

Croissant, J. Engendering technology: culture, gender, and work. *Knowledge and Society*, 12, 189–207, 2000.

McGaw, J. Reconceiving technology: why feminine technologies matter, in *Archaeology and Gender*, R.P. Wright, Ed. University of Pennsylvania Press, Philadelphia, 1996, pp. 52–75.

McIlwee, J.S. and Robinson J.G. *Women in Engineering: Gender, Power, and Workplace Culture*. SUNY Press, Albany, 1992.

Oldenziel, R. *Making Technology Masculine: Men, Women, and Modern Machines in America, 1870–1945*. Amsterdam University Press, Amsterdam, 1999.

Pacey, A. *The Culture of Technology*. Blackwell, Oxford, 1983.

Ranson, G. and Reeves, W. J. Gender, earnings, and proportions of women: lessons from a high-tech occupation. *Gender and Society*, 10, 2, 168–184, 1996.

Roos, P.A. and Gatta M.L. The gender gap in earnings: trends, explanations, and prospects, in *Handbook of Gender and Work*, Powell, G., Ed. Sage, Thousand Oaks, CA, 1999, pp. 95–125.

Stanley, A. *Mothers and Daughters of Invention: Notes for a Revised History of Technology*. Scarecrow Press, Metuchen, NJ, 1993.

Vare, E.A. and Ptacek, G. *Mothers of Invention: From the Bra to the Bomb : Forgotten Women & their Unforgettable Ideas*. Morrow, New York, 1988.

Vare, E.A. and Ptacek, G. *Patently Female: From AZT to TV Dinners: Stories of Women Inventors and their Breakthrough Ideas*. Wiley, New York, 2002.

Witz, A. *Professions and Patriarchy*. Routledge, London, 1992.

Gene Therapy

In 1971, Australian Nobel laureate Sir F. MacFarlane Burnet thought that gene therapy (introducing genes into body tissue, usually to treat an inherited genetic disorder) looked more and more like a case of the emperor's new clothes. Ethical issues aside, he believed that practical considerations forestalled possibilities for any beneficial gene strategy, then or probably ever. Bluntly, he wrote: "little further advance can be expected from laboratory science in the handling of 'intrinsic' types of disability and disease." Joshua Lederberg and Edward Tatum, 1958 Nobel laureates, theorized in the 1960s that genes might be altered or replaced using viral vectors to treat human diseases. Stanfield Rogers, working from the Oak Ridge National Laboratory in 1970, had tried but failed to cure argininemia (a genetic disorder of the urea cycle that causes neurological damage in the form of mental retardation, seizures, and eventually death) in two German girls using Swope papilloma virus. Martin Cline at the University of California in Los Angeles, made the second failed attempt a decade later. He tried to correct the bone marrow cells of two beta-thalassemia patients, one in Israel and the other in Italy. What Cline's failure revealed, however, was that many researchers who condemned his trial as unethical were by then working toward similar goals and targeting different diseases with various delivery methods. While Burnet's pessimism finally proved to be wrong, progress in gene therapy was much slower than antibiotic or anticancer chemotherapy developments over the same period of time.

In September 1990, National Institutes of Health investigators Michael Blaese, French Anderson, and Kenneth Culver carried out the first authorized gene therapy clinical trial. Their patient, 4-year-old Ashanthi DeSilva, suffered from a form of severe combined immune deficiency (SCID) caused by lack of an essential enzyme, adenosine deaminase (ADA). The protocol called for a recombinant virus—beginning with a denatured mouse leukemia virus, into which a human gene for adenosine deaminase and promoter/marker genes were spliced—to correct rather than destroy her circulating white blood cells. A blood specimen was taken from the child, and her cells were grown in the presence of the virus and later reinfused. In this and several subsequent cases, SCID patients also received a bovine-derived adenosine deaminase. Results were encouraging, but the addition of a replacement enzyme clouded proof that gene therapy alone was responsible for improving the patient's health. In April 2000, Alain Fischer at the Hôpital Necker Enfants-Malades in Paris reported success following a similar strategy in treating five patients with a related type of SCID for which no replacement enzyme existed. Only then was there unchallenged proof for the principle of gene therapy.

During the 1990s, however, researchers in the U.S., Europe, and Japan conducted over 400 clinical gene therapy trials, with the largest group targeting cancers. Some strategies aimed at stimulating immune responses to tumor antigens; others sought to insert a functioning tumor suppressor gene, or, alternatively, to inhibit gene function to stop tumor growth. One protocol attempted to change the nature of a deadly brain cancer by inserting a retrovirus recombined with the thymidine kinase gene from the cold sore virus, herpes simplex. Because brain cells in the adult human grow very slowly if at all and retroviruses can only infect dividing cells, researchers believed they could alter only the tumor's cells. After surgery to implant virus-producing cells and allowing time for the virus to spread, the antiviral drug ganciclovir was administered to create a lethal toxin through its interaction with thymidine kinase. While some patients apparently benefited from this approach and adverse reactions were not significant, positive indications were hardly compelling, other than suggesting a need for more research.

As clinical experiments continued, more limitations became apparent. Several protocols relied on adenoviruses as vectors to deliver a functional gene to the lung tissue of cystic fibrosis patients, but the normal immune response rejected this approach in much the same way it fights a cold. No viral vector could be targeted, thus gene delivery carried a risk of disrupting some other essential genetic function. In the cancer trials, tumor cell heterogenicity and mutability defeated any stunning breakthrough. Many patients in these studies died, though in only one instance at the University of Pennsylvania in 1999 was the therapy itself responsible for a death. Ironically, had gene therapy proved to be more effective, it may also have been less safe.

In the late 1990s, investigators began pursuing another approach—gene repair. One protocol was directed at treating Crigler-Nijjar syndrome, a liver enzyme deficiency responsible for a fatal bilirubin clearance disorder. The approach relied on a synthesized oligonucleotide that intentionally mismatched the point mutation of the disease on chromosome 2. The aim was to provoke normal gene repair enzymes to correct the problem in

enough cells (perhaps no more than 5 percent of the liver) so that a life-saving repair would result. The technique used liposomes for gene delivery. Because gene repair relied on nonviral delivery and aimed at generating self-repair in cells, its potential for efficacy and safety exceeded that of gene therapy. However, as of early 2002 no clinical trials had begun.

While gene therapy had limited success, it nevertheless remained an active area for research, particularly because the Human Genome Project, begun in 1990, had resulted in a "rough draft" of all human genes by 2001, and was completed in 2003. Gene mapping created the means for analyzing the expression patterns of hundreds of genes involved in biological pathways and for identifying single nucleotide polymorphisms (SNPs) that have diagnostic and therapeutic potential for treating specific diseases in individuals. In the future, gene therapies may prove effective at protecting patients from adverse drug reactions or changing the biochemical nature of a person's disease. They may also target blood vessel formation in order to prevent heart disease or blindness due to macular degeneration or diabetic retinopathy. One of the oldest ideas for use of gene therapy is to produce anticancer vaccines. One method involves inserting a granulocyte-macrophage colony-stimulating factor gene into prostate tumor cells removed in surgery. The cells then are irradiated to prevent any further cancer and injected back into the same patient to initiate an immune response against any remaining metastases. Whether or not such developments become a major treatment modality, no one now believes, as MacFarland Burnet did in 1970, that gene therapy science has reached an end in its potential to advance health.

See also **Genetic Engineering, Applications; Genetic Engineering, Methods**

G. Terry Sharrer

Further Reading

Culver, K.W. *Gene Therapy: A Primer for Physicians*, 2nd edn. Mary Ann Liebert, Larchmont, NY, 1996.

Friedmann, T. *The Development of Human Gene Therapy*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999.

Krniec, E.B. Gene therapy. *American Scientist*, May 1999.

Lyon, J. and Gorner, P. *Altered Fates: Gene Therapy and the Retooling of Human Life*. W.W. Norton, New York, 1995.

Porter, R. *The Greatest Benefit to Mankind: A Medical History of Humanity*. W.W. Norton, New York, 1997.

Verma, I.M. and Somia, N. Gene therapy—promises, problems and prospects. *Nature*, 389, 239–242, 1997.

Walters, L. and Palmer, J.G. *The Ethics of Human Gene Therapy*. Oxford University Press, Oxford, 1997.

Useful Websites

Human Genome Project information, gene therapy: http://www.ornl.gov/hgmis/medicne/genetherapy.html

Genetic Engineering, Applications

For centuries, if not millennia, techniques have been employed to alter the genetic characteristics of animals and plants to enhance specifically desired traits. In a great many cases, breeds with which we are most familiar bear little resemblance to the wild varieties from which they are derived. Canine breeds, for instance, have been selectively tailored to changing esthetic tastes over many years, altering their appearance, behavior and temperament. Many of the species used in farming reflect long-term alterations to enhance meat, milk, and fleece yields. Likewise, in the case of agricultural varieties, hybridization and selective breeding have resulted in crops that are adapted to specific production conditions and regional demands.

Genetic engineering differs from these traditional methods of plant and animal breeding in some very important respects. First, genes from one organism can be extracted and recombined with those of another (using recombinant DNA, or rDNA, technology) without either organism having to be of the same species. Second, removing the requirement for species reproductive compatibility, new genetic combinations can be produced in a much more highly accelerated way than before. Since the development of the first rDNA organism by Stanley Cohen and Herbert Boyer in 1973, a number of techniques have been found to produce highly novel products derived from transgenic plants and animals.

At the same time, there has been an ongoing and ferocious political debate over the environmental and health risks to humans of genetically altered species. The rise of genetic engineering may be characterized by developments during the last three decades of the twentieth century.

1970 to 1979

The term genetic engineering was probably first coined by Edward L. Tatum during his 1963 Nobel Prize acceptance speech, but it was not until the following decade that many of the potential applications of gene transfer became more apparent to the emerging field of molecular biology. However, the period was witness to a clash

between the new molecular biologists who were pioneering the techniques of genetic engineering and their more industrially related colleagues in microbiology. Whereas the microbiologists, with their roots in the fermentation industries, were more prepared to recognize the potential of genetic engineering, molecular biologists were initially much more concerned with the threats of environmental risk.

As a consequence of a letter to *Science* and *Nature* in 1974, known as the Berg letter after its first signatory Paul Berg, molecular biologists instigated a voluntary moratorium on gene transfer work until such time as the community was satisfied that necessary safety measures and procedures had been put in place. Two further key events were instrumental in moving the field forward. The first was the Asilomar Conference in February 1975, which addressed safety measures and, to a lesser extent, prospects for future applications. The second was the publication in 1996 of the first guidelines for gene transfer research released by the U.S. National Institutes for Health, which effectively lifted the moratorium. In 1977, Genentech reported the manufacture of the human hormone somatostatin in bacteria genetically engineered to contain a synthetic gene that produced a human protein. This step is widely considered to represent the opening moments of modern biotechnology production.

1980 to 1989

The early 1980s is the first period in which large-scale investments were made in the biotechnology industry against the anticipation of huge profits to follow. Indeed, the first biotechnology stocks to be floated on the markets in the 1980s rose in value far more rapidly than had any other sector up to that time. In 1980 the U.S. Supreme Court allowed patent protection for a genetically modified "oil-eating" bacterium, providing powerful financial incentives for biotechnology companies to expand research. The year 1980 also saw the introduction of the polymerase chain reaction (PCR) technique, through which DNA sequences are multiplied many times *in vitro*, which became the foundation of much of the work to follow. In 1981 a team at Ohio University produced the first transgenic animals, mice in this case. Shortly afterward, Harvard University released details of studies in which mice were engineered to carry a human gene that increased susceptibility to a form of human cancer. The "OncoMouse," as it became known, could be used to test the carcinogenicity of different compounds and as a model for developing cures for cancer. The OncoMouse, for which Harvard filed a patent in 1984, focused much of the ensuing debate on the future health implications of genetic engineering. By 1983, the first patents had been granted on genetically engineered plants (actually only for the use of an antibiotic resistance "marker gene," that allowed researchers to select transformed plants by their ability to survive exposure to an otherwise lethal dose of an antibiotic). In 1985 the first U.S. field trials of genetically engineered crops (tomatoes with a gene for insect resistance and tobacco plants with herbicide resistance) took place, and in 1986, genetically engineered tobacco plants, modified with addition of a gene from the bacterium *Bacillus thuringiensis* (Bt) to produce a insecticidal toxin, making the hybrid resistant to the European corn borer and other pests, underwent field trials in the U.S. and France.

1990 to 2000

The 1990s saw considerable growth in a wide range and variety of biotechnological applications, though without necessarily fulfilling the huge

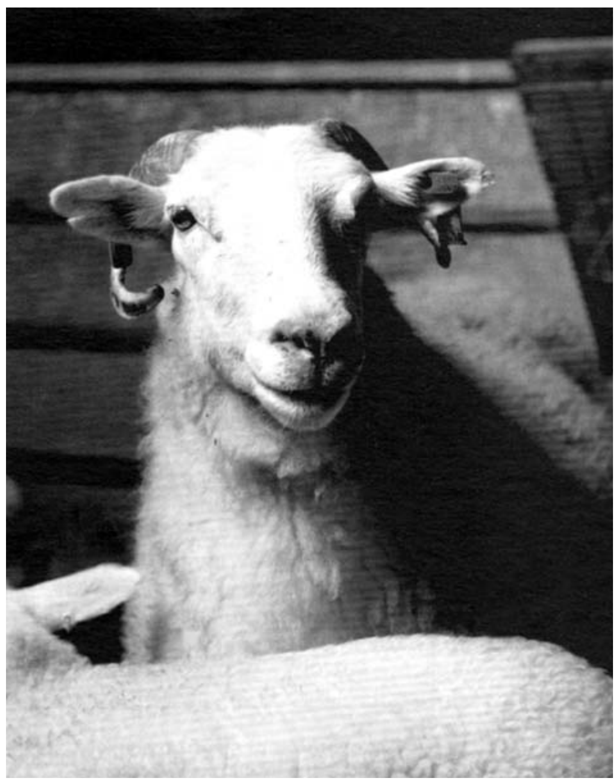


Figure 1. Tracy—PPL Therapeutics' transgenic ewe, modified in 1992 to produce a human protein in her milk. [*Courtesy of PPL Therapeutics.*]

expectations evident in the early 1980s. In respect to animal biotechnology products, a number of events can be seen to have defined the decade.

Sizeable resources were directed at the production of proteins and drug compounds in transgenic animals, resulting in over 50 varieties of genetically modified (GM) bioreactors. These methods had a number of advantages over traditional cell culture production including higher production volumes, particularly in respect to those proteins (such as human albumin) that cannot be produced in a sufficient volume using other available techniques. On a considerably smaller scale, research in the 1990s also focused on the production of transgenic animals as sources of transplant tissues and organs. However, the decade closed with little progress seen in either reducing tissue rejection or overcoming anxieties about transpecies disease.

By far the largest research activity was within the field of plant biotechnology. For instance, genetically modified herbicide-tolerant (GMHT) crops were intended to enable varieties to withstand chemical treatments that would normally damage them. The same concept was applied to the production of insect and virus-resistant plant varieties, in addition to altering the way fruits ripen so that they can withstand increased storage and travel stresses. In 1994, the Flavr Savr tomato, designed to delay ripening and resist rotting, became the first whole genetically engineered food to be approved for sale in the U.S. (China commercialized virus-resistant tobacco plants in the early 1990s). In 2000, a rice variety was genetically engineered to contain a gene that increases the vitamin A content of the grains. Similar improvements could be made to the composition of other important food staples. Another widespread application of genetic engineering prevented plants from pollinating in order to limit the chances of cross-fertilization with other species. A more controversial aspect of genetically modified plants was their inability to reproduce so that growers would be unable to collect seeds for replanting, and thus forced to purchase seed from the supplier each season.

The 1990s were also characterized by what became known as the "GM debate." Although the strength of the controversy varied considerably throughout the world, with much greater intensity in Europe than in the U.S., anxieties continued to focus on a number of potentially adverse environmental effects arising from GM foods. First, there were concerns that GM crops will indirectly reduce wild plant biodiversity through intensification of industrial agriculture (increased use of pesticides and herbicides) and potentially threaten species higher up the food chain, for example invertebrates that feed on the weeds, and their bird and mammal predators. Second, the debate has focused on the risk of cross-pollination and gene transfer between GM and non-GM plants. Contamination has implications for the labeling of foods, for organic methods, and for seed production. Finally, more intensive production methods, together with enhanced transportation tolerance, were seen to exacerbate other environmental problems, particularly pollution and global warming.

See also **Biotechnology, Breeding, Animal Genetic Methods; Breeding, Plant Genetic Methods; Gene Therapy; Genetic Engineering, Methods; Genetic Screening and Testing; Pesticides**

NIK BROWN

Further Reading

Boylan, M. and Brown, K.E. *Genetic Engineering: Science and Ethics on the New Frontier*. Prentice Hall, Upper Saddle River, NJ, 2001.

Bud, R. Molecular biology and the long-term history of biotechnology, in *Selection Indices and Prediction of Genetic Merit in Animal Breeding*, Cameron, N.D., Ed. CAB International, Oxford, 1997.

Reeve, E., Ed. *Encyclopedia of Genetics*. Fitzroy Dearborn, London and Chicago, 2000. See entries on Rice genetics: engineering vitamin A; Pharmaceutical proteins from milk of transgenic animals.

Thackray, A., Ed. *Private Science: Biotechnology and the Rise of the Molecular Sciences*. University of Pennsylvania Press, Philadelphia.

Useful Websites

Advisory Committee on Novel Foods and Processes. Annual Report 2000, FSA/0013/0301: http://www.foodstandards.gov.uk/multimedia/pdfs/acnfp2000

Genetic Engineering, Methods

The term "genetic engineering" describes molecular biology techniques that allow geneticists to analyze and manipulate deoxyribonucleic acid (DNA). At the close of the twentieth century, genetic engineering promised to revolutionize many industries, including microbial biotechnology, agriculture, and medicine. It also sparked controversy over potential health and ecological hazards due to the unprecedented ability to bypass traditional biological reproduction. This article describes common genetic engineering techniques, excluding gene therapy and organism cloning, which are covered in separate entries in this encyclopedia.

Recombinant DNA (rDNA) technology involves combining DNA from different organisms, often from radically different species. In 1970, Howard Temin and David Baltimore independently isolated an enzyme, reverse transcriptase, which cuts DNA molecules at specific sites. Paul Berg, who was awarded the Nobel Prize in Chemistry in 1980 for his work with DNA, in 1972 used a similar restriction enzyme, ligase, to paste two DNA strands together to form a recombinant DNA molecule. In 1973, Stanley Cohen at Stanford University and Herbert Boyer at the University of California in San Francisco, combined two plasmids (short, circular pieces of single-stranded DNA) that naturally exist in many bacteria. Each contained a different antibiotic resistance gene. The new ''recombinant'' plasmid was inserted into a bacterial cell, and the cell produced both types of antibiotic resistance proteins. This bacteria was the first recombinant DNA organism.

Rapid advances occurred in the late 1970s, including splicing genes from higher organisms such as humans into bacterial plasmids. By the early 1980s valuable proteins such as insulin, interferon, and human growth hormone were being synthesized using recombinant bacteria hosts (such as *Escherichia coli*), and there was anticipation that these proteins would soon be produced on an industrial scale.

During meiosis, the production of gametes from dividing sperm and egg cells, the two copies of each chromosome exchange DNA by a process called ''genetic crossover.'' A crossover between two regions of DNA causes them to be separated and no longer co-inherited. Thus the frequency of co-inheritance is an indirect measure of the distance between DNA regions. This ''genetic distance'' is measured in ''Morgans,'' after Thomas Hunt Morgan's studies of the phenomenon in fruit flies c.1915. The determination of genetic distance is called ''genetic mapping.''

Genetic mapping in humans was relatively uncommon prior to rDNA technology. In 1978, Yuet Wai Kan at the University of California, San Francisco, discovered a region of DNA genetically close to the sickle cell anemia gene using genetic mapping techniques. Genetic mapping in humans then expanded rapidly.

In 1975 Ed Southern at Oxford University in the U.K. invented the ''Southern Blot.'' The procedure relies on two other technologies: gel electrophoresis and DNA probes. Gel electrophoresis was developed in the 1950s for separating molecules of different sizes on a gel subjected to an electric current. DNA probes, small strands of radioactive DNA with a known base sequence, bind to DNA with a base pair sequence that matches that of the probe, and the binding can be visualized using X-ray film. Using these two techniques, DNA strands containing a desired sequence can be isolated from a mixture of strands (see Figure 2).

These technologies can be used to produce a ''physical map,'' a collection of overlapping DNA fragments arranged in their proper order. Two DNA fragments overlap if a probe that recognizes a unique sequence hybridizes to both fragments. A physical map is a useful tool when searching for genes. If a gene has been genetically mapped to a specific region of DNA, a physical map of this region can be used to locate and isolate this gene.

In 1977, methods for sequencing DNA were developed in two separate locations, by Walter Gilbert at Harvard University and by Fred Sanger at Cambridge University. These DNA sequencing processes were automated in the mid-1980s and marketed by Applied Biosystems, DuPont, and other companies. The ''Sanger method'' is the most common method; it uses dideoxynucleotides (ddNTPs), which are similar to deoxynucleotides (dNTPs), the normal components of DNA, but they lack an oxygen atom. There are four forms, corresponding to the four dNTP types found in DNA: adenine (A), thymine (T), guanine (G), and cytosine (C). When any of these is incorporated into a synthesizing DNA strand, DNA synthesis stops. Sequencing involves four separate DNA synthesis reactions, each containing one of the four ddNTPs and each using the original DNA strand as a template. DNA synthesis occurs in each reaction tube and ends when ddNTP gets incorporated into the strand. If the reaction is allowed to continue long enough, each of the four tubes will have a full assortment of DNA strands of varying

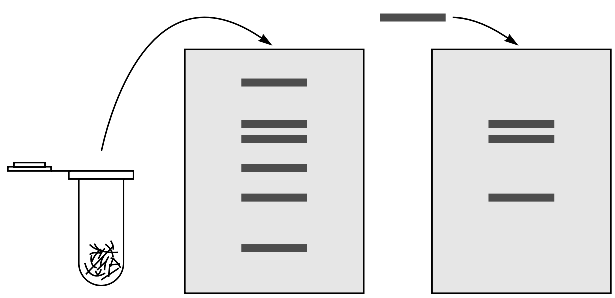


Figure 2. Southern Blot.
[*Source: Dr. John Medina, of the University of Washington, Seattle, Washington. Used with permission from Brinton, K. and Lieberman, K.-A. Basics of DNA Fingerprinting, University of Washington, 1994.*]

lengths, all ending with the ddNTP present in the reaction tube. When these strands are separated by gel electrophoresis, the sequence of the original DNA molecule can be determined by simply reading the sequencing gel from bottom to top (see Figure 3).

In 1985 Kary Mullis, a technician with the Cetus Corporation in the U.S., invented the polymerase chain reaction (PCR), a process that could rapidly identify and replicate a segment of DNA with a known sequence out of a complex mixture of DNA. The technique was time saving and simple compared with the Southern blot and was used for a wide variety of applications, including genetic and physical mapping. Mullis received the 1993 Nobel Prize in Chemistry for this technique.

PCR involves successive rounds of DNA synthesis of double-stranded DNA (dsDNA). The two strands are first separated and used as a template to synthesize another strand each; these strands then form a new dsDNA. Both the original and the new dsDNA are then subjected to this same process. This cycle is performed from 20 to 40 times, and the result is an exponential amplification of the DNA of interest (see Figure 4).

Site-directed mutagenesis, invented in 1978 by Michael Smith (1932–2000) of the University of British Columbia in Canada, is a targeted alteration of DNA sequence. It involves inserting DNA to be mutated into a plasmid and exposing it to a short DNA fragment containing one changed base. This fragment then binds to the insert. A new DNA molecule is synthesized from this DNA fragment, and this results in a double-stranded plasmid. Multiplying this plasmid in bacteria will produce equal numbers of single-stranded plasmids with and without the mutation; the mutated DNA can then be isolated (see Figure 5). In 1982 Smith managed to produce the protein products of these induced mutations, which allowed for analysis of protein function. Smith was the co-recipient of the 1993 Nobel Prize in Chemistry for this technique and shared the prize with Mullis.

Introducing a recombinant DNA molecule into the host cell varies according to the host organism. Bacteria can take up DNA molecules naturally. In animal cells in culture and fertilized egg cells, DNA can be injected directly into the cell

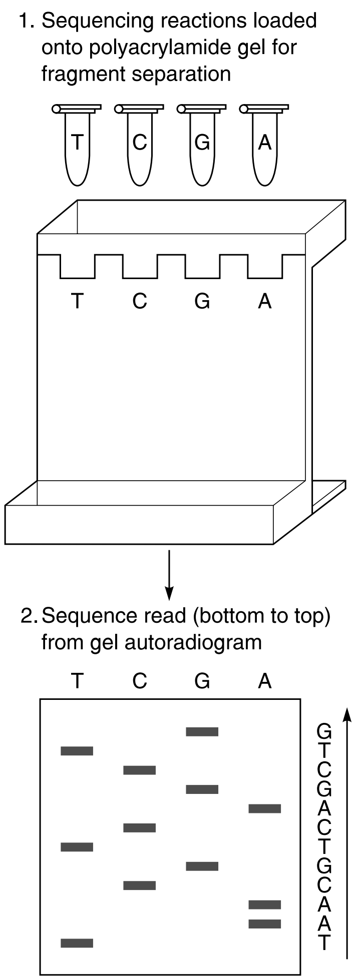


Figure 3. Sanger sequencing. Once the four dideoxynucleotide sequencing reactions are concluded, each reaction tube, containing one of the four dideoxynucleotides (T, C, G, or A) are subjected to gel electrophoresis, which separates the different synthesized strands. The sequence of the original DNA strand is then ascertained by reading the gel from bottom to top. [*Source: Reprinted with permission from U.S. Department of Energy Human Genome Program, Primer on Molecular Genetics, 1992.*]

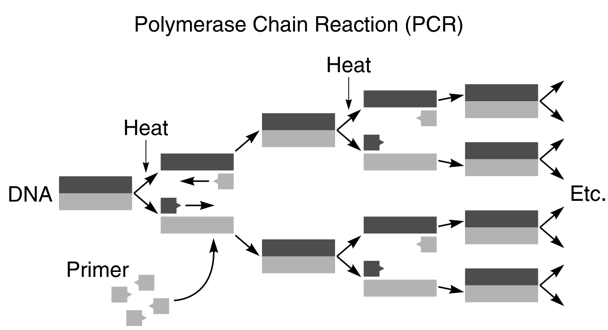


Figure 4. Exponential amplification of DNA using the polymerase chain reaction (PCR). Heat is used to separate the DNA strands. Primers are short DNA strands used to initiate DNA synthesis. [*Source: Reprinted with permission from Wrobel, S. Serendipity, science, and a new hantavirus. FASEB Journal, 9, 1247–1254, 1995.*]

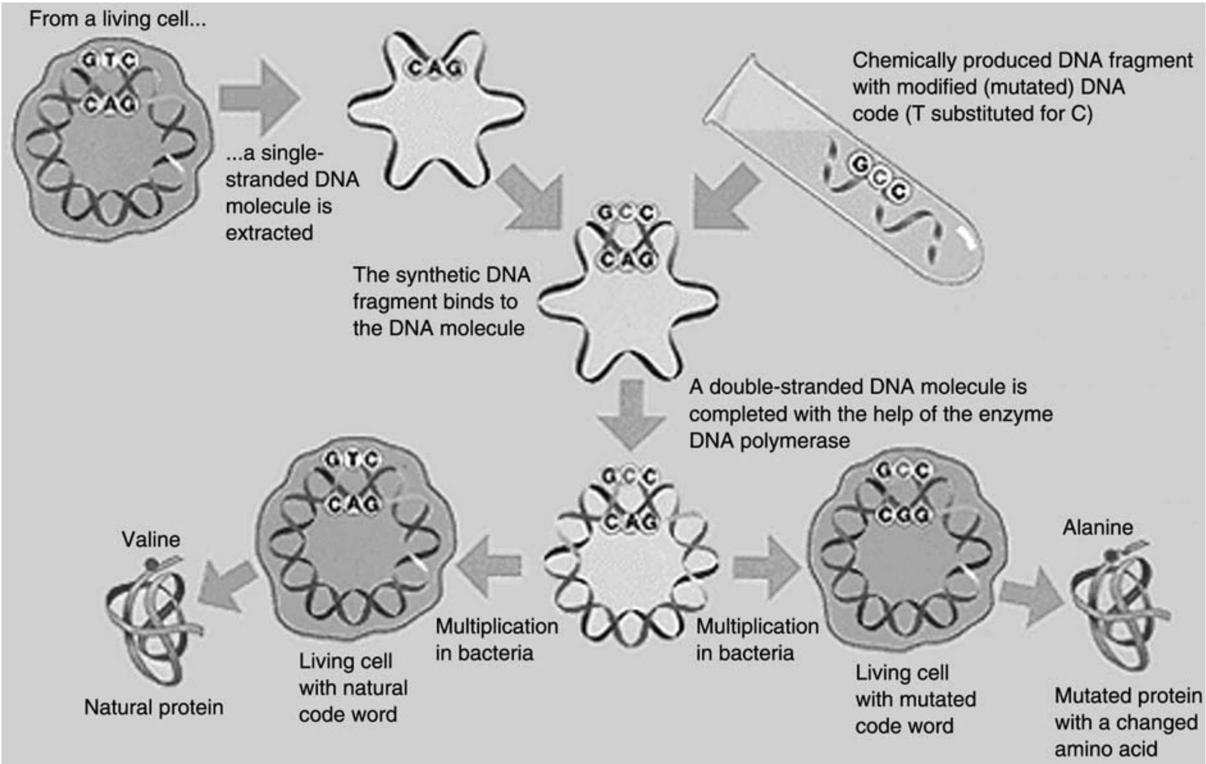


Figure 5. Site-directed mutagenesis.
[*Source: Reprinted with permission from the Royal Swedish Academy of Sciences from the Nobel E-Museum, Illustrated Presentation: The Nobel Prize in Chemistry, 1993*.]

nucleus, where it integrates with the host DNA. In plants, destruction of the cell wall permits DNA to be readily taken up by the plant cell, which then spontaneously reforms the cell wall. Alternatively, a vector or small carrier molecule can be derived from plasmids, bacteriophages, or plant and animal viruses. In 1977 it was shown that part of the tumor-inducing (Ti) plasmid of the soil-borne bacterial plant pathogen *Agrobacterium tumefaciens* (which causes galls or tumors on the plant near soil level) is transferred to the host plant cell during infection. If the tumor gene is removed and replaced with another gene, the bacteria can introduce the beneficial gene into plants. The Ti plasmid is now widely used as a vector system for transferring foreign genes into the plant genome.

More commonly, a retroviral vector that has been rDNA modified can be used to infect mammal cells: the viral DNA is integrated into the host cells.

Many modifications to these common techniques have increased their effectiveness and usefulness. Together, they allowed scientists to manipulate DNA in unprecedented ways and created major topics of interest and research for industry, government, and the general public.

See also **Cloning, Testing and Treatment Methods; Gene Therapy; Genetic Engineering, Applications**

TED EVERSON

Further Reading

Aldridge, S. *The Thread of Life: The Story of Genes and Genetic Engineering*. Cambridge University Press, New York, 1996.
Cantor, C.R. and Smith, C.L. *Genomics: The Science and Technology Behind the Human Genome Project*. Wiley, New York, 1999.
Cook-Deegan, R. *The Gene Wars: Science, Politics, and the Human Genome*. W.W. Norton, New York, 1996.
Fox Keller, E. *The Century of the Gene*. Harvard University Press, Cambridge, MA, 2000.
Green, J.J. and Rao, V.B., Eds. *Recombinant DNA Principles and Methodologies*. Marcel Dekker, New York, 1998.
Sheldon K. *Biotechnics and Society: The Rise of Industrial Genetics*. Praeger, New York, 1991.
Watson, J.D., Gilman, M., Witkowski, J. and Zoller, M. *Recombinant DNA: A Short Course*, W.H. Freeman, San Francisco, 1983.

Wright, S. Recombinant DNA technology and its social transformation, 1972–1982. *Osiris*, 2, 303–360, 1986.

Useful Websites

Southern blotting: http://www.biology.washington.edu/fingerprint/blot.html
Human Genome Project information: http://www.ornl.gov/hgmis
Nobel Prize in Chemistry, 1993: http://www.nobel.se/chemistry/laureates/1993/illpres/index.html

Genetic Screening and Testing

The menu of genetic screening and testing technologies now available in most developed countries increased rapidly in the closing years of the twentieth century. These technologies emerged within the context of rapidly changing social and legal contexts with regard to the medicalization of pregnancy and birth and the legalization of abortion. The earliest genetic screening tests detected inborn errors of metabolism and sex-linked disorders. Technological innovations in genomic mapping and DNA sequencing, together with an explosion in research on the genetic basis of disease which culminated in the Human Genome Project (HGP), led to a range of genetic screening and testing for diseases traditionally recognized as genetic in origin and for susceptibility to more common diseases such as certain types of familial cancer, cardiac conditions, and neurological disorders among others. Tests were also useful for forensic, or nonmedical, purposes.

The earliest genetic testing procedures began in the 1950s when amniocentesis was used to obtain fetal cells from amniotic fluid for analysis to ascertain the sex of fetuses being carried by women with a family history of sex-linked diseases such as hemophilia. However, this procedure was not rapidly adopted, not in the least part because of the nearly universal illegality of abortion until the late 1960s. A more precise form of genetic testing using amniocentesis emerged in the mid-1960s. The procedure followed the discovery in the late 1950s that the common form of the condition known then as mongolism and now Down's syndrome was caused by the presence of three copies of chromosome 21 (instead of the normal two). The condition came to be called trisomy 21 as other major chromosomal abnormalities associated with disease conditions were identified. Although it was initially difficult to detect abnormalities using cultures of fetal cells obtained via amniocentesis, these technological problems were overcome by the mid-1960s. Testing of pregnant women both due to family history but more widely to "advanced maternal age" (known to be associated with increased risk of trisomy 21) became common, particularly following the legalization of abortion and later with increased litigation against physicians, particularly in the U.S.

Also in the 1960s, widespread testing of newborns for the metabolic disease phenylketonuria (PKU), which results from inheritance of two copies of a mutatedgene responsible for making the enzyme phenylalanine hydroxylase, was made possible through the development of basic and economically efficient screening technologies. Newborn screening for PKU and other metabolic diseases involves taking a blood sample from the baby's heel and placing small blood spots on a piece of filter paper from which samples are punched out to test for metabolic errors. This method was developed by the U.S. microbiologist Robert Guthrie, whose name is now associated with the filter paper samples known as "Guthrie cards." Guthrie successfully lobbied U.S. state governments to require testing of all newborns, which he argued would reduce public health costs by allowing early dietary intervention and perhaps avoiding the disease's damaging early effects. The testing was not uncontroversial, however, since the diet was difficult to maintain and the rate of incorrect diagnoses (false positives and false negatives) was initially high.

A particularly notable early genetic screening program occurred in the U.S. in the early 1970s for sickle cell anemia. This condition affects approximately 1 in 400–600 African–Americans who have two copies of the sickle cell version of the β-hemoglobin gene, and 1 in 10–12 African–Americans are carriers for this gene. Mandatory screening programs were implemented, focusing on school-aged children and young adults, despite limited treatment options. These programs were widely criticized for problems with the accuracy of the test results, limited community consultation, and the stigmatization and discriminatory effects that often resulted, particularly for those who were carriers but who did not in fact have sickle cell anemia.

Few additional testing options became available until the 1980s and 1990s, when genetic technologies began to allow screening for many more disease conditions. The range of conditions detectable through prenatal testing expanded to include not only trisomy 21 and other major chromosomal abnormalities, but also conditions associated with point gene mutations such as cystic fibrosis. Screening procedures were initially targeted at

those couples with a family history of the particular genetic disease and members of subpopulations where it was recognized that certain diseases were more common (e.g., testing for Tay–Sachs disease among those of Ashkenazi Jewish backgrounds). A third targeted group were women who had abnormal test results early in their pregnancies from procedures such as ultrasound or screening for maternal serum alpha-fetoprotein (MSAFP), a simple blood test that allows early detection of increased risk of neural tube defects and Down's syndrome. Prenatal testing is performed using amniocentesis or chorionic villus sampling (CVS), a newer technology that allows detection of genetic anomalies at an earlier stage of pregnancy (i.e., during the first trimester, or first three months). Although originally developed in the 1970s, CVS did not become commonly used until the 1980s. It was considered more desirable than amniocentesis by many consumers since it allows earlier decisions about pregnancy termination, although the procedure carries greater risks.

Adult presymptomatic testing also began for a limited range of inherited conditions, notably the late-onset neurological disorder Huntington's disease, which is a dominant condition transmitted via one copy of the affected gene so that every individual with one affected gene expresses the disorder (compared to recessive disorders, where two copies of the affected gene are needed). Many of these genetic testing programs, especially those associated with adult onset disorders, have not been popular for a number of reasons: lack of availability of treatment options, anticipated problems with discrimination with regard to life and health insurance, employment and other areas, concerns about predictive value (i.e., the lack of availability of information about age of onset, severity of disease), and the potential for possibly revealing the genetic status of other family members such as parents or twins.

The Human Genome Project, begun in 1990 and completed in 2003, refined existing mapping and sequencing technologies and provided more information about the function of genes and genetic pathways, resulting in an increased number of diseases that can be detected using genetic testing and screening. In some cases, genetic testing can now provide predictive information about the likely medical effects of particular genetic mutations or DNA sequence variations known as single nucleotide polymorphisms (SNPs), although for many diseases this information remains limited. Screening for predispositions to develop serious conditions such as certain types of familial cancer, neurological conditions, and cardiac disease also has become available, and such programs undoubtedly will expand to include many more common conditions.

Consideration of the ethical, legal, and social implications of clinical genetic testing raises concerns about privacy, potential misuse, and counseling. If genetic information is disclosed, how can we prevent discrimination by health insurers and employers? Knowledge of genetic differences may lead to anxiety about disease, relationship breakdown, and social stigmatization. A positive result for a predictive test for adult-onset diseases such as Huntington's disease does not guarantee a cure or treatment, and individuals in high-risk families may decide not to be tested. Screening for susceptibility to diseases is also likely to benefit few people in the developing world.

Genetic screening techniques are now available in conjunction with *in vitro* fertilization and other types of reproductive technologies, allowing the screening of fertilized embryos for certain genetic mutations before selection for implantation. At present selection is purely on disease grounds and selection for other traits (e.g., for eye or hair color, intelligence, height) cannot yet be done, though there are concerns for eugenics and "designer babies." Screening is available for an increasing number of metabolic diseases through tandem mass spectrometry, which uses less blood per test, allows testing for many conditions simultaneously, and has a very low false-positive rate as compared to conventional Guthrie testing. Finally, genetic technologies are being used in the judicial domain for determination of paternity, often associated with child support claims, and for forensic purposes in cases where DNA material is available for testing.

See also **Diagnostic Screening; Fertility, Human; Gene Therapy; Genetic Engineering, Applications; Genetic Engineering, Methods**

Rachel A. Ankeny

Further Reading

Andrews, L. *Future Perfect: Confronting Decisions about Genetics*. Columbia University Press, New York, 2001.

Cowan, Ruth Schwartz. Aspects of the history of prenatal diagnosis. *Fetal Diagn. Ther.*, 8, 1, 10–17, 1993.

Holtzman, N.A. *Proceed with Caution: Predicting Genetic Risks in the Recombinant DNA Era*. Johns Hopkins University Press, Baltimore, 1989.

Kevles, D.J. and Hood, L., Eds. *Code of Codes: Scientific and Social Issues in the Human Genome Project*. Harvard University Press, Cambridge, 1992.

Lippman, A. Prenatal genetic testing and screening: constructing needs and reinforcing inequities. *Am. J. Law Med.*, 17, 15–50, 1991.
Lynch, M. and Jasanoff, S., Eds. Contested identities: science, law and forensic practice. *Soc. Stud. Sci.* 28, 5–6, 1998.
Millington, D.S. Newborn screening for metabolic diseases. *American Scientist*, 90, 1, 40–47, 2002.
Markel, H. The stigma of disease: implications of genetic screening. *Am. J. Med.*, 93, 2, 209–215, 1992.
Rothman, B.K. *The Tentative Pregnancy: Prenatal Diagnosis and the Future of Motherhood*. Viking, New York, 1986.
U.S. National Research Council. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington D.C., 1996.

Useful Websites

Paul, D.B. The history of newborn phenylketonuria screening in the U.S. in *Promoting Safe and Effective Genetic Testing in the United States: Final Report of the Task Force on Genetic Testing*, Holtzman, N.A. and Watson M.S., Eds.
Markel, H. Scientific advances and social risks: historical perspectives of genetic screening programs for sickle cell disease, tay-sachs disease, neural tube defects and Down syndrome, 1970–1997, in *Promoting Safe and Effective Genetic Testing in the United States: Final Report of the Task Force on Genetic Testing*, Holtzman, N.A. and Watson M.S., Eds.
Both available at http://www.nhgri.nih.gov/ELSI/TFGT_final

Global Positioning System (GPS)

The use of radio signals for navigation began in the early twentieth century. Between 1912 and 1915, Reginald Fessenden of Boston, Massachusetts, transmitted radio waves from designated shore-based stations to correct ship chronometers and give mariners a sense of direction. The prospect of greater accuracy came in 1940 when Alfred Loomis of the National Defense Research Council suggested simultaneous transmission of pulsed signals from a pair of precisely surveyed ground stations to ship receivers and use of the difference in arrival times of the two signals to calculate a line of position. Developed during World War II by the Radiation Laboratory at the Massachusetts Institute of Technology, the system Loomis had proposed became known as Loran (LOng RAnge Navigation), and it supported convoys crossing the Atlantic. Subsequent improvements significantly increased Loran accuracy, but the technology was still limited to two dimensions—latitude and longitude. Not until 1960 did Ivan Getting of Raytheon Corporation in Lexington, Massachusetts, propose the first three-dimensional type of Loran system. Intended to solve navigational problems associated with rail-based, mobile intercontinental ballistic missiles, Raytheon's system was called Mosaic (MObile System for Accurate ICBM Control). That system was never developed, however, because a new technology—artificial satellites—provided the basis for more precise, line-of-sight radio navigation.

The NAVSTAR (NAVigation System Timing And Ranging) Global Positioning System (GPS) provides an unlimited number of military and civilian users worldwide with continuous, highly accurate data on their position in four dimensions—latitude, longitude, altitude, and time—through all weather conditions. It includes space, control, and user segments (Figure 6). A constellation of 24 satellites in 10,900 nautical miles, nearly circular orbits—six orbital planes, equally spaced 60 degrees apart, inclined approximately 55 degrees relative to the equator, and each with four equidistant satellites—transmits microwave signals in two different L-band frequencies. From any point on earth, between five and eight satellites are "visible" to the user. Synchronized, extremely precise atomic clocks—rubidium and cesium—aboard the satellites render the constellation semiautonomous by alleviating the need to continuously control the satellites from the ground. The control segment consists of a master facility at Schriever Air Force Base, Colorado, and a global network of automated stations. It passively tracks the entire constellation and, via an S-band uplink, periodically sends updated orbital and clock data to each satellite to ensure that navigation signals received by users remain accurate. Finally, GPS users—on land, at sea, in the air or space—rely on commercially produced receivers to convert satellite signals into position, time, and velocity estimates.

Drawing the best concepts and technology from several predecessor navigation systems, engineers synthesized one that became known in the early 1970s as GPS. Those previous efforts began with Transit or the Naval Navigation Satellite System (NNSS), developed in the late 1950s by the Applied Physics Laboratory at Johns Hopkins University in Baltimore, Maryland, and used operationally by both the U.S. Navy and commercial mariners from 1962 to 1996. Also contributing to the mix was the U.S. Naval Research Laboratory's Timation satellite project that began during 1964 to provide very precise timing and time transfer between points on Earth. A third element originated with Aerospace Corporation's Project 57, which included aircraft navigation using satellite signals and which the

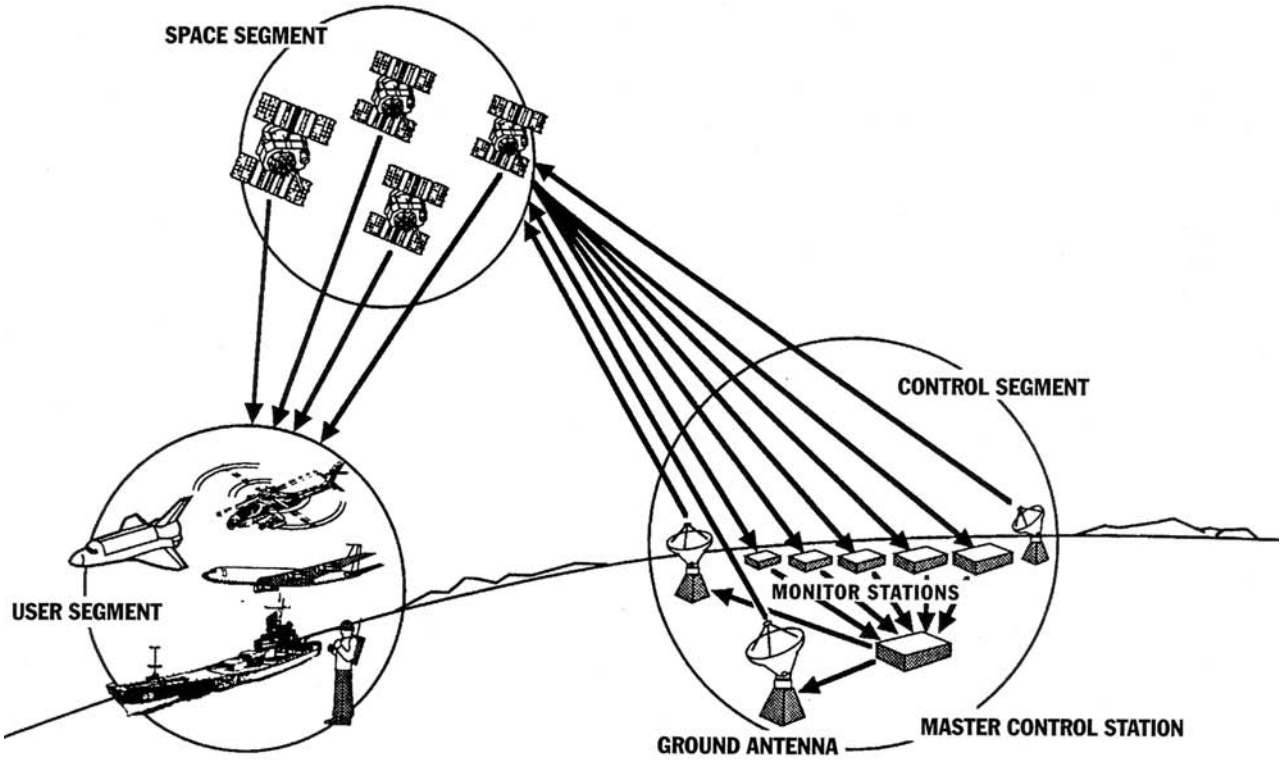


Figure 6. Major Global Position System (GPS) segments. [*Courtesy U.S. Air Force.*]

U.S. Air Force, under direction from the Advanced Research Projects Agency (ARPA), pursued as Project 621B beginning in October 1963. The formation in 1973 of a multiservice or joint program office headed by a U.S. Air Force colonel, Bradford Parkinson, drew together these various programmatic threads led to approval from the Defense System Acquisition and Review Council (DSARC) to proceed with GPS development. This led quickly to the launch of two Navigation Technology Satellites to explore technical capabilities. By the time the first Block I or Navigation Development Satellite was launched in February 1978, an initial control segment was functioning and several kinds of user equipment were being tested at Yuma Proving Ground on the Arizona–California border.

The GPS space segment evolved steadily from the late 1970s into the 1990s. Between February 1978 and October 1985, the U.S. Air Force successfully launched and operated ten of eleven Block I satellites (Figure 7). In February 1989, the first Block II fully operational GPS satellite went into orbit, and near the end of 1993, both military and civilian authorities had declared achievement of initial operational capability. With the launch of a Block IIA satellite on March 9, 1994, a fully operational, 24-satellite GPS constellation was completed. To sustain the space segment, the U.S. Air Force began launching Block IIR satellites on July 23, 1997. The latter also enhanced performance through greater autonomy and increased protection against radiation—natural or man-made. Further improvements scheduled for 2005 with Block IIF and sometime after 2009 with Block III would make GPS less vulnerable to jamming by delivering a more powerful signal to military users. In addition, these future GPS satellites would be able to surge signals over a specific area by using spot beams and would employ new frequencies for civil "safety of life" applications.

Designed for use by both military and civilian parties at no charge, GPS originally offered two levels of service—precise and standard. With specially equipped receivers, authorized users could pinpoint their precise position and time with at least 22-meter horizontal, 27.7-meter vertical, and 200-nanosecond accuracy. Through a Selective Availability (SA) feature, the U.S. Defense Department could intentionally degrade GPS signals for the majority of users, who possessed standard receivers. The latter were accurate only to within 100 meters horizontally,

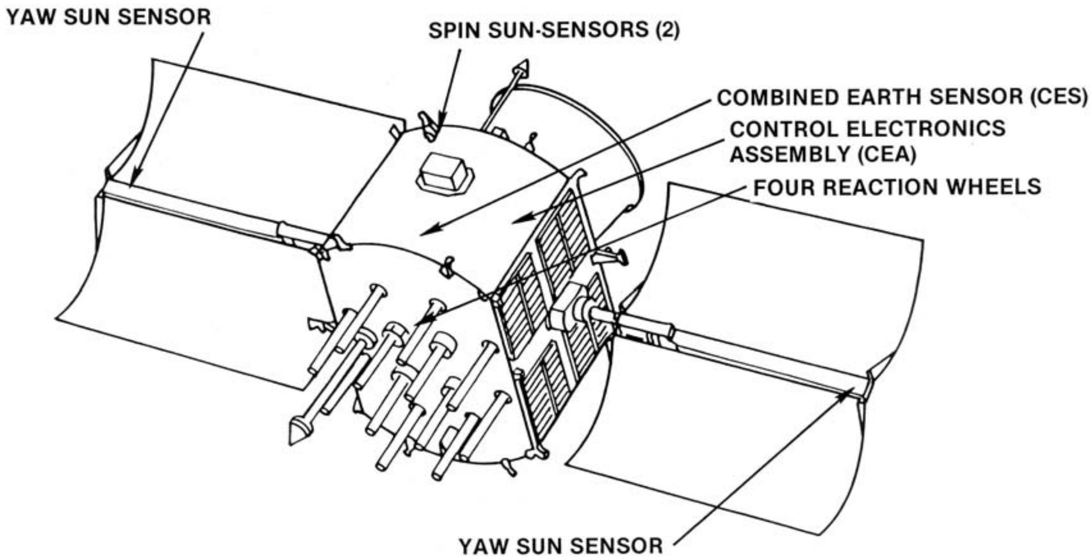


Figure 7. Global Positioning System Block I satellite diagram indicating avionics equipment. [*Source: Courtesy U.S. Air Force*.]

156 meters vertically, and 340 nanoseconds. By 1996, however, it had become evident that the exploding market for GPS among civilians demanded greater precision. Consequently, on May 1, 2000, President Clinton announced immediate discontinuation of SA. By giving everyone access to the more precise level of service, his administration hoped to encourage private sector investment in what was already a multibillion-dollar growth industry.

The uses for GPS data have become mind boggling in number and variety, as have the number of GPS devices. On the military side, the uses include aerial refueling, rendezvous operations, forward air control, location and recovery of downed pilots, targeting and delivery of precision weapons, and computer network integration. Civilians employ GPS to manage natural resources, control maritime and land-based traffic, survey and map terrain, conduct search and rescue operations, monitor earthquakes and storms, improve athletic performance, and link communication networks and disparate databases. In automobile GPS devices alone, there was a tenfold increase from 1,100,000 units in 1996 to 11,300,000 worldwide in 2001. As GPS receivers became smaller and cheaper, they could be found in everything from cellular telephones to wristwatches.

At the beginning of the twenty-first century, GPS signals were even being used for navigation in space, and it seemed likely that aviation authorities would soon rely on an augmented GPS capability for air traffic control. Nonetheless, skeptics doubted that GPS would someday govern people's lives to the extent enthusiasts predicted, and alarmists warned that jamming or hacking could render a GPS-dependent world militarily and economically helpless. Supporters worried about maintaining stable funding sources for the GPS space and control segments, but critics in the U.S. Congress accused the U.S. Air Force of poor program execution as well as budgeting too far ahead of what it needed simply to sustain the system. Meanwhile, engineers pondered future GPS designs that might improve positional accuracy to 30 centimeters and time transfer to 1 nanosecond before the end of the first decade of the new century.

See also **Gyrocompass and Inertial Guidance; Radionavigation**

RICK W. STURDEVANT

Further Reading

Adams, Thomas K. GPS vulnerabilities. *Mil. Rev.*, 81, 10–16, 2001.

Getting, I.A. The global positioning system. *IEEE Spectrum*, 30, 36–47, 1993.

Marquis, W. M is for modernization: Block IIr–m satellites improve on a classic. *GPS World*, 12, 36–42, 2001.

National Academy of Public Administration and National Research Council. *The Global Positioning System: Charting the Future*. National Academy Press, Washington D.C., 1995.

National Research Council. *The Global Positioning System: A Shared National Asset*, National Academy Press, Washington D.C., 1995.

Nordwall, B.D. Using pseudo-satellites to foil GPS jamming. *Aviation Week & Space Technology*, 155, 54–61, 2001.

Parkinson, B.W., and Spilker, J.J. Jr., Eds. *Global Positioning System: Theory and Applications*, 2 vols. American Institute of Aeronautics and Astronautics, Washington D.C., 1996.

Rip, M.R. and Hasik, J.M. *The Precision Revolution: GPS and Future of Aerial Warfare*. Naval Institute Press, Annapolis, 2002.

Useful Websites

Adams, T.K. GPS vulnerabilities. *Military Review*: http://www-cgsc.army.mil/milrev/English/MarApr01/adams.htm

Dana, P.H. *Global Positioning System Overview*: http://www.colorado.edu/geography/gcraft/notes/gps/gps.html

Federal Aviation Administration (FAA) Global Positioning System Product Team: http://gps.faa.gov/

Franson, P. GPS's Sweet Dreams: http://www.upside.com/texis/mvm/story?id=369123830

NAVSTAR GPS Joint Program Office (SMC/CZ): http://gps.losangeles.af.mil

Navtech, GPS Related Internet Sites: http://www.navtechgps.com/links.asp

Globalization

Technology was pivotal to globalization throughout the twentieth century. While the transfer and diffusion of technological innovations fueled economic growth, ever-greater global economic interdependence was enabled by new transport and communication technologies, which progressively tamed time and distance over the course of the century. Technology was thus both a direct agent of the economic growth and development underpinning globalization and, simultaneously, a medium of globalization facilitating the ever-greater movement of materials, people, knowledge, and ideas on which economic interdependence depended. Daniele Archibugi and Jonathan Michie (1997) distinguished between the former concentration upon the *globalization of technology* and the latter concentration on *the technologies of globalization*. These two categories provide the focus for this entry. A common thread is provided by the mutual significance of the institutional, organizational, and broader socio-cultural systems of which technologies are part, as they are crucial to understanding the dynamics of their development, dissemination, and utilization. Railways, for example, not only facilitated the movement of people and goods but also inspired innovations in business organization, were intimately involved in the development of the national identities of various nations, and crucially shaped conceptions of time and timekeeping. While the post-World War II period of intense global economic growth and integration is most widely identified with the term globalization, analogous activity earlier in the century.

The Technologies of Globalization

The dawn of the twentieth century was preceded by some three decades of unprecedented and intense integration and consolidation in the world economy brought about by the traffic in materials, people, and ideas that resulted from railroads, steamships, and the telegraph. Newly emerging technologies including the telephone, the wireless, and the internal combustion engine were all soon destined to bring the world even closer together. Ongoing industrialization both facilitated and drove these developments with the early industrial, most notably northern European, nations relying heavily on this traffic to supply an escalating demand for both basic foodstuffs and industrial raw materials. This increasing demand stimulated a search for less expensive, and thus more remote, sources of supply and a concomitant outflow of capital, skilled labor, and technology to develop these resources. A number of countries, notably Argentina, Australia, Canada, New Zealand, South Africa, and Uruguay, experienced rapid development at this time as a result of the influx of overseas capital, labor, and technology resulting from this demand for their primary products. Perhaps the greatest single factor facilitating these developments was the improvement in transport, witnessed by increased investment in port facilities and railways both within industrialized nations and primary exporting countries. The development of these latter regions was, however, highly reliant on the diffusion of technological improvements, particularly in the mining and agricultural sectors.

The pre-World War I period paralleled the globalization that marked the closing decades of the twentieth century in a number of other ways. The average size of firms and business enterprises grew rapidly, sound forms of business organization increasingly became the norm, and dynamic multinational business activity emerged on a large scale. Changes in trade, manufacturing, and technology, such as the increasing use of electric motors, also resulted in the first significant cases of painful deindustrialization. Parts of northern England, for example, which were first to industrialize with

steam and textiles, could not respond rapidly enough to these changes, and long-lasting pools of unemployment were created. The limit of this analogy to the closing decades of the century is underlined by the fact that when this period was abruptly halted by World War I, it was largely only Western Europe, North America, and Japan that had industrialized. Much of the rest of the globe, however, including some of the poorest nations such as those of Africa, were increasingly drawn upon as sources of raw materials.

The period after World War I saw the emergence of automobiles and air travel, which both vastly accelerated the taming of time and distance and extended these changes to far more people and further corners of the globe. However foreign trade declined in importance relative to the prewar period. As a result some contemporary economists argued that technological progress and the spread of industrialization would contribute to declining trade between nations. These arguments were never broadly accepted, and the surge in economic growth and accompanying expansion in trade that followed World War II effectively ended the discussion. The 1920s and 1930s did, nonetheless, lay a foundation for the coming surge in growth and globalization. The Middle East developed rapidly as a source of oil on the back of western investment. Russia's swift industrialization was facilitated by burgeoning domestic resources, while Japan set the scene for later developments by developing a regional economic hinterland, most notably in Korea.

In the years after 1945, there was not only development of the intimate global economic interdependence that inspired the term globalization but also a parallel and ultimately perhaps more profound political and socio-cultural interdependence. The most commonly perceived and perhaps most extensive of these latter developments was the global ascendency of American popular culture, the pervasiveness of which profoundly marked late twentieth century life. On the political front, the first truly global institutions came with the founding of the United Nations in 1945 and the development of the Bretton-Woods institutions (the World Bank and International Monetary Fund) that played particularly critical roles in late twentieth century economic globalization. Communications and reinforcing developments in entertainment technologies were key to the advance of American popular culture. The rise of television, the consolidation of Hollywood as a global focus for film making, and the emergence of rock and roll music and the youth culture that accompanied it paralleled radical geopolitical changes external to the U.S. Such changes as the rise of welfare states in Europe, decolonization, and postwar rebuilding, left a vacuum that American popular culture was destined to fill. Technology played both an instrumental and a profoundly cultural role in these developments. Many technologies including automobiles and motorbikes, and for some politicians hydroelectric dams and nuclear power stations, took on an iconic status that was close to totemic. Analogous developments occurred on many fronts, from pop music and movies to fast food, soft drinks, and personal entertainment equipment and telephones.

The technological character of World War II itself signaled a fundamental shift in the importance and role assigned technology that would be borne out in increased research and development (R&D) spending in the postwar period. The development of radar, rockets, digital computers and the harnessing of science to technological imperatives, pioneered by the U.S. Manhattan Project to develop the first atomic bomb, all raised immense possibilities that were soon to be the focus of frenetic activity. R&D spending in the U.S. rose from 1 percent of gross domestic product (GDP) in 1950 to 2.8 percent in 1960. In the USSR over the same period, it grew from 1.2 percent to 2.5 percent, a trend reflected in most other leading industrial nations. By 1993 the U.S. figure was still at 2.8 percent, while other leading investors such as Germany at 2.5 percent, the U.K. at 2.15 percent, and South Korea at 2.1 percent reflected similar priorities and an increasing acknowledgement of the role of science and technology in wealth creation. The vast increases in trade accompanying these changes brought a significant growth in foreign direct investment (FDI) by which commercial interests in one country invest in activities in another. While U.S. multinational corporations initially dominated this growth, European multinationals were featured in the investment flows from the 1970s, with Japanese multinationals emerging during the 1980s. By the mid-1990s many other countries including Hong Kong, Korea, Singapore, Taiwan, China, Mexico, and Brazil were involved.

Railway construction continued in many developing regions after 1945, but in most industrialized countries, railways declined in significance in the face of competition from road transport and the growing importance of aviation. This led to rising oil consumption, reflected in a significant increase in the amount of oil traded globally and in oil tanker capacity. Increased air travel further pro-

moted greater commercial interdependence by facilitating the face-to-face contact invaluable in business. From the late 1950s containerization vastly increased the speed and efficiency of the handling and transportation of goods, while emerging satellite links would revolutionize intercontinental telecommunications.

Much of the increase in economic interdependence after 1945 reflected an intensification of earlier trends and developments. For example, a significant element in Japan's rise to economic prominence was her skill in harnessing innovation and technological systems, notably in manufacturing and transport, in which innovation was as much organizational as technical. The success of Japan's postwar strategy to gain access to raw materials emulated the earlier success of the U.S. and northern Europe by harnessing improved technologies in shipping and bulk cargo handling across great distances and between countries. Throughout southern Asia and Australasia in particular, export-oriented primary commodity production and transportation systems were developed in which the size and scale of all elements were integrated to optimize efficiencies and economies of scale. Such elements included extractive technologies such as mining operations, transportation to port, port and shipping facilities, and Japanese domestic port and manufacturing operations. The success of Japanese innovation and production management in this period is highlighted by the way Japan's Ministry for International Trade and Industry (MITI) methods, such as "lean production" in the automobile industry, were so widely emulated by other leading industrial nations during the final decades of the century.

Technology was increasingly regarded as a key to competitive advantage, placing an ever-greater emphasis on information and knowledge and on their management rather than on technology *per se*. While this emphasis reflected the economic rationalism prevalent at this time, which closely resembled the *laissez-faire* perspective of the late nineteenth and early twentieth century, it was also an acknowledgment of the contemporary significance of knowledge in product innovation and development. This shift was reflected both in policy and in the nature of some of the most significant technologies of the closing decades of the century. The microprocessor-based information and communication technologies and genetic biotechnologies that emerged at this time were both centered on the encryption, transmission, and manipulation of knowledge and information.

The development of microprocessors in the 1970s unleashed unprecedented information processing and communication capabilities that became particularly marked with an exponential growth in their power and in the extent of the Internet in the 1980s. These technologies greatly facilitated both the economic deregulation of the leading industrial economies (given particular impetus by President Ronald Reagan in the U.S. and Prime Minister Margaret Thatcher in the U.K.) and the surge in transnational commerce and related rise of speculative global capital markets that were such a dominant feature of the late twentieth century. For example, the deregulation of electricity industries, a major concern of this period, was made possible by new systems able to coordinate transactions between numerous generators and consumers, scheduling and administering the transmission and the quality of supply. The scale of trading in speculative global capital markets facilitated by these technologies, including trading of such major economic indicators as currencies, was so great that it challenged the volume of the markets in material goods and industrial output by the close of the century.

These technologies soon permeated everyday life in the industrial world; electronic domestic monetary transactions were common and by the end of the century home computer use for work, entertainment, and information retrieval purposes were widespread. The significance of these changes was the subject of great debate as the century closed. What is clear is that there were significant differences with the scale and pace of past developments. A particularly notable example is the Internet, which achieved ubiquity within about a decade, many times faster than earlier comparable technologies such as the telephone. By the end of the century the management, transmission, and manipulation of information had become both a mainstay of wealth creation and of significance for lifestyles more generally. This was quite distinct from an earlier narrower emphasis on the production of material goods and from earlier "knowledge economies" in which stock market activity was tied directly to the production of material goods.

A similar emphasis on the manipulation of information is central to genetic biotechnology, which also emerged and developed rapidly in this period. The potential of this technology, a major product of some of the larger contemporary multinational agrichemical and pharmaceutical companies, was still largely untested and controversial as the century closed. However the implica-

tions of human understanding and intervention in the "coding of life" were widely viewed as being at least as profound as those of the information and communication technologies discussed above.

The "dematerialization" evident in these new information focused technologies and trends and the weakening of links to tradition and place characteristic of late twentieth century globalization were understood in a variety of ways. Some interpreted these changes in terms of the advent of a new postindustrial or postmodern society. However, while these notions flagged the significant changes, if not transformations, evident in leading industrial economies, some of the poorest nations, including many African ones, remained underdeveloped and poorly integrated with the global economy and broader global culture even by the close of the century.

Some of the most thoughtful insights into the nature of the changes marking late twentieth century life were made by those with a limited understanding of technology but a profound understanding of society and its dynamics. Anthony Giddens, a leading twentieth century social theorist, described how the implications of these technologies of globalization can be understood in terms of an increasing distancing of time and space involving a growing separation of social activity from localized contexts. The implications he draws from this include a novel requirement for trust in complex large-scale sociotechnical systems such as airlines or the Internet, a mounting ability to apply updated knowledge to personal circumstances that reinforces the dynamic complexity of contemporary life, and the way differentials in knowledge increasingly map into differentials in power. Giddens' analysis serves to underline how technology had become less a tool for human ends and more an integral element of human life and broader culture over the course of the twentieth century.

The Globalization of Technology

While late twentieth century globalization was both an effect and a cause of the large-scale development, exploitation, distribution, and transfer of scientific and technological innovation, mainstream economists such as Rosenberg (1982) regarded technology as a "black box" for most of the century. In this view, technology was simply a "public good" autonomous of broader economic and social factors. From the 1980s however, the economics of technology became more widely acknowledged and by the 1990s increasingly studied by policy makers. Inspired particularly by Karl Marx and Joseph Schumpeter, the economics of technology concentrates on innovation as the key to technological change and economic growth. It encompasses a variety of approaches and ascribes a varying importance to firms, forces at the national level, global forces, or to particular technological systems or regimes. However, perhaps surprisingly in light of a widespread perception of the late twentieth century power of multinational corporations, the concept of "national systems of innovation" was particularly influential. Echoing the insights of the twentieth century German economist Friedrich List, these ideas underline the critical importance of education and training, investment in R&D, national integration and infrastructure, and the extent and quality of interaction between the many national players involved in innovation.

The increasing internationalization of business, witnessed by the twentieth century surge in FDI, underlined the fact that national technological competence was regarded as a key to harnessing technology. While direct national investment in R&D has the potential to benefit offshore interests in a globalized world, national technological competence was seen as more likely to encourage local innovation, inward investment, and an ability to utilize technologies developed elsewhere. Also, while education and training are crucial to national technological competence, the structuring of economic incentives and broader aspects of national culture such as attitudes to technology are also significant elements. The globalization of technology was thus regarded as being as much contingent on context and culture as it was on hard cash or engineering.

However, while business and governments universally promoted late twentieth century globalization as an incontestable good, in practice it advanced the interests of the most powerful industrial countries and did little to redress contemporary patterns of inequity and unequal development. These are complex and controversial matters in which technology played a small but not insignificant part. Technology transfer was generally acknowledged to be essential for more equitable economic growth and development, but commercial imperatives significantly constrained such transfer. For example, resistance by developed countries to effective technology transfer in the environmental domain, dating back to the very first multilateral environmental summit in Stockholm in 1972, was significant in impeding negotiation and agreement in both the Montreal Protocol (1987) and Kyoto Protocol (1997) pro-

cesses. Another, rather different, example of the role of technology in these matters is provided by the speculative global capital markets. Widely implicated in the Asian "financial crisis" of 1997–1998, these markets were a target for regulation and reform as the century closed.

The complex and increasingly global relationships of technology and societies over the course of the twentieth century had many ramifications. Not only was the globe effectively a smaller, more closely knit place by the end of the century, but many aspects of life in the most developed parts of the globe were thoroughly technology-driven in ways quite distinct from earlier societies. In the industrialized world, technology formed part of the fabric of life in previously unimaginable ways. Electricity networks, airlines, the Internet, and highway systems fashioned what people did in the late twentieth century far more intensively than railroads, steamships, and the telegraph had done at the start of the century. Technological developments increasingly informed cultural developments. By the end of the century, digital technology was central to both the form and content of many types of visual art, music, and communication, with the latter opening up new forms of interpersonal and communal interactions that were evolving rapidly as the century closed.

While the quality of life enabled by these changes was immeasurably better in the developed countries at the end of the century, some developments, such as those in the environmental arena, were profoundly disturbing. What was becoming clearer as the century closed was that a thoroughly technologized globe would only be viable under conditions that addressed the complex interaction of technology and society. Although the twentieth century afforded the insight that this problem was as much a political, cultural, and economic matter as it was a technical one, it would be for the next century to unravel the repercussions of this.

See also **Communications; Computers, Uses and Consequences; Organization of Technology and Science; Research and Development in the Twentieth Century; Social and Political Determinants of Technological Change; Technology, Society, and the Environment; Telecommunications; Transport**

STEPHEN HEALY

Further Reading

Archibugi, D. and Michie, J. Technological globalization and national systems of innovation: an introduction, in *Technology, Globalization and Economic Performance*, Archibugi, D. and Michie, J., Eds. Cambridge University Press, Cambridge, 1997.

Archibugi, D. and Michie, J., Eds. *Trade, Growth and Technical Change*. Cambridge University Press, Cambridge, 1998.

Bunker, S. and Ciccantell, P. Restructuring space, time, and competitive advantage in the capitalist world economy: Japan and raw materials transport after World War II, in *A New World Order? Global Transformations in the Late Twentieth Century*, Smith, D. and Borocz, J., Eds. Greenwood Press, Westport, CT, 1995.

Cowan, R.S. *A Social History of American Technology*. Oxford University Press, New York, 1997.

Fichman, M. *Science, Technology and Society: A Historical Perspective*. Kendall/Hunt Publishing, Dubuque, IA, 1993.

Giddens, A. *The Consequences of Modernity*. Polity Press, Cambridge, 1990.

Kenwood, G. and Lougheed, A.L. *The Growth of the International Economy 1820–2000: An Introductory Text*. Routledge, London, 1999.

McMillan, C. Shifting technological paradigms: from the US to Japan, in *States Against Markets: The Limits of Globalization*, Boyer, R. and Drache, D., Eds. Routledge, New York, 1996.

Rosenberg, N. *Inside the Black Box: Technology and Economics*. Cambridge University Press, Cambridge, 1982.

Talalay, M., Farrands, C. and Tooze, R., Eds. *Technology, Culture and Competitiveness: Change and the World Political Economy*. Routledge, London, 1997.

Green Chemistry

The term "green chemistry," coined in 1991 by Paul T. Anastas, is defined as "the design of chemical products and processes that reduce or eliminate the use and generation of hazardous substances." This voluntary, nonregulatory approach to the protection of human health and the environment was a significant departure from the traditional methods previously used. While historically people tried to minimize exposure to chemicals, green chemistry emphasizes the design and creation of chemicals so that they do not possess intrinsic hazard.

Within the definition of green chemistry, the word chemical refers to all materials and matter. Therefore the application of green chemistry can affect all types of products, as well as the processes to make or use these products. Green chemistry has been applied to a wide range of industrial and consumer goods, including paints and dyes, fertilizers, pesticides, plastics, medicines, electronics, dry cleaning, energy generation, and water purification.

A fundamental aspect of green chemistry is the recognition that intrinsic hazard is simply another property of a chemical substance. Because proper-

ties of chemicals are caused by their molecular structure, the properties can be altered or modified by changing the chemical structure. Chemists and molecular scientists have been developing ways of manipulating chemical structure since the nineteenth century, and green chemistry uses the same expertise to design the property of minimal hazard into the molecular structure of chemicals.

The types of hazards that can be addressed by green chemistry vary and can include physical hazards such as explosiveness and flammability, toxicity including carcinogenicity (cancer-causing) and acute lethality, or global hazards such as climate change or stratospheric ozone depletion. Therefore, in the same way that a substance can be designed to be colored blue or to be flexible, it can also be designed to be nontoxic.

Principles of Green Chemistry

Paul T. Anastas and John E Warner's twelve principles of green chemistry outline a framework for chemists and chemical engineers to use in the quest to design more environmentally benign products and processes. These principles look at the entire life cycle of a product or process from the origins of the materials that go into its manufacture to the ultimate fate of the materials after they have finished their useful life. Through the use of the principles of green chemistry, scientists have been able to reduce the impact of chemicals in the environment. The principles are as follows:

1. It is better to prevent waste than to treat or clean up waste after it is formed.
2. Synthetic methodologies should be designed to maximize the incorporation of all materials used in the process into the final product.
3. Wherever practicable, synthetic methodologies should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
4. Chemical products should be designed to achieve efficacy of function while reducing toxicity.
5. The use of auxiliary substances (e.g., solvents, separation agents, etc.) should be made unnecessary wherever possible and, innocuous when used.
6. Energy requirements should be recognized for their environmental and economic impacts and should be minimized. Synthetic methods should be conducted at ambient temperature and pressure.
7. A raw material or feedstock should be renewable rather than depleting wherever technically and economically practicable.
8. Unnecessary derivatization (e.g., blocking group, protection/deprotection, temporary modification of physical/chemical properties) should be avoided whenever possible.
9. Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
10. Chemical products should be designed so that at the end of their function they do not persist in the environment and break down into innocuous degradation products.
11. Analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.
12. Substances and the form of a substance used in a chemical process should be chosen so as to minimize the potential for chemical accidents, including releases, explosions, and fires.

Green Chemistry Research and Development

Research and development in the field of green chemistry occurs in several different areas.

Alternative Feedstocks Historically, many of the materials that went into making the products we used often depleted finite resources, such as petroleum, or were toxic. Green chemistry is active in developing ways of making the products that we need from renewable and non-hazardous substances such as plants and agricultural wastes.

Benign Manufacturing Synthetic methods used to make chemical materials have often involved the use of toxic chemicals such as cyanide and chlorine. In addition, these methods have at times generated large quantities of hazardous and toxic wastes. Green chemistry research has developed new ways to make these synthetic methods more efficient and to minimize wastes while also ensuring that the chemicals used and generated by these methods are as nonhazardous as possible.

Designing Safer Chemicals Once it is ensured that the feedstocks and the methods to make a substance are environmentally benign, it is important to make certain that the end product is as nontoxic as possible. By understanding what makes a product harmful at the molecular level (as in the field of molecular toxicology), chemists

and molecular scientists can design the molecular structure so that it cannot cause this harm.

Green Analytical Chemistry The detection, measurement, and monitoring of chemicals in the environment through analytical chemistry has been a part of the environmental movement since its beginning. Instead of measuring environmental problems after they occur, however, green chemistry seeks to prevent the formation of toxic substances and thus prevent problems before they arise. By putting sensors and instruments into industrial manufacturing processes, green analytical chemistry is able to detect trace amounts of a toxic substance and to adjust process controls to minimize or stop its formation altogether. In addition, while the traditional methods of analytical chemistry used substances such as hazardous solvents, green analytical methods are being developed to minimize the use and generation of these substances while carrying out the analysis.

The Need for Green Chemistry

The need for green chemistry has evolved with the increased use of chemicals. Many companies around the world, including those within the U.S., China, Germany, Italy, Japan, and the U.K., are adopting green chemistry, and there are several reasons for this. First, green chemistry can effectively reduce the adverse impact of chemicals on human health and the environment. The second reason is economic; many companies have found that it is less expensive, and even profitable, to meet environmental goals using green chemistry. The monetary savings come from the combined factors of higher efficiency, less waste, better product quality, and reduced liability. The third reason is legal. Many environmental laws and regulations target hazardous chemicals, and following all of the requirements can be a long and complex process. By using green chemistry, companies are finding they are able to comply with the law in much simpler and cheaper ways. Finally, green chemistry is a fundamental science-based approach. This means that because it deals with the problem of hazard at the molecular level, green chemistry approaches can be applied to all kinds of environmental issues.

Conclusion

Since 1991, there have been many advances in green chemistry in academic research and in industrial implementation. These advances, however, represent just a small fraction of the potential applications of green chemistry. Because the products and processes that form the basis of the economy and infrastructure are based on the design and utilization of chemicals and materials, the challenges facing green chemistry are vast.

See also **Biotechnology; Chemicals; Feedstocks; Environmental Monitoring**

PAUL T. ANASTAS

Further Reading

Anastas, P.T. and Farris, C.A., Eds. *Benign by Design: Alternative Synthetic Design for Pollution Prevention*. Oxford University Press, New York, 1994.

Anastas, P.T., Heine, L.G. and Williamson, T.C., Eds. *Green Chemical Syntheses and Processes*. Oxford University Press, New York, 2000.

Anastas, P.T., Heine, L.G. and Williamson, T.C., Eds. *Green Engineering*. ACS Symposium Series No. 766, Oxford University Press, New York, 2001.

Anastas, P.T. and Williamson, T.C., Eds. *Green Chemistry: Designing Chemistry for the Environment*. Oxford University Press, New York, 1996.

Anastas, P.T. and Williamson, T.C., Eds. *Green Chemistry: Frontiers in Benign Chemical Synthesis and Processes*. 1998.

Lankey, R.L. and Anastas, P.T., Eds. *Advancing Sustainability through Green Chemistry and Engineering*. ACS Symposium Series No. 823, Oxford University Press, New York, 2002.

Poliakoff, M.and Anastas, P. A principled stance. *Nature*, 413, 257–258, 2001.

Useful Websites

U.S. Environmental Protection Agency, Green Chemistry Program: http://www.epa.gov/greenchemistry/

Green Chemistry Institute at the American Chemical Society: http://www.acs.org/greenchemistryinstitute, 2002.

Gyrocompass and Inertial Guidance

Before the twentieth century, navigation at sea employed two complementary methods, astronomical and dead reckoning. The former involved direct measurements of celestial phenomena to ascertain position, while the latter required continuous monitoring of a ship's course, speed, and distance run. New navigational technology was required not only for iron ships in which traditional compasses required correction, but for aircraft and submarines in which magnetic compasses cannot be used. Owing to their rapid motion, aircraft presented challenges for near instantaneous navigation data collection and reduction. Electronics furnished the exploitation of radio and the adaptation of a gyroscope to direction finding

through the invention of the nonmagnetic gyrocompass.

Jean Bernard Léon Foucault's invention of the gyroscope in the mid-nineteenth century spurred the idea of a spinning rotor within two gimbals aligned with north that might be adapted to direction finding. In 1908 Hermann Anschütz-Kaempfe introduced a gyrocompass for naval use, intending it for submarine guidance to reach the North Pole. His gyrocompass was installed in a German warship in 1910. Meanwhile, the American Elmer Ambrose Sperry, who began work on a seaworthy gyrocompass in 1896, established the Sperry Gyroscope Company in 1910 and installed his version in a warship in 1911. World War I furnished the first rigorous test of the gyrocompass. During the war years, gyroscopes enabled the installation of automatic pilots in ships and aircraft. After the war, Sperry Gyroscope Company became Sperry Corporation (and much later the Unisys Corporation), the leading world manufacturer and developer of gyrocompasses and other gyroscopic devices. Numerous American, German, and British inventors contributed modifications and improvements, although many of them remained unknown outside their industry because of military secrecy. Patent infringement claims have further obscured the lineage of gyrocompasses. Anschütz-Kaempfe's vision was not realized until 1958 with the voyages of the American submarines *USS Nautilus* and *Skate* under the polar ice, navigated by inertial guidance, the most important adaptation of the gyro in the last half of the century.

The gyrocompass relies on a key property of a gyro: its rotor or flywheel, mounted within gimbals, when set in motion can be oriented such that its axis points in any direction. The utility of a gyro used as a compass is its stability. Once oriented to north, the rotor axis maintains the plane of rotation unless acted upon by an external influence (known as gyroscopic inertia) and is unaffected by magnetism. Various influences, however, cause the axis in motion to precess, meaning that the rotation axis responds at right angles to the influence of an applied force. Gyrocompasses may wander due to the earth's rotation or when the gyro-controlled vehicle is stationary. A vehicle's motion may also influence the direction of the rotor axis (tilt and drift are common terms for axial wander). North is not assumed automatically by the rotor's axis.

Considerable sophistication in the application of gyroscopes to navigation had occurred before World War II. Wiley Post's 1933 solo air circumnavigation involved multiple gyro applications including a Sperry autopilot. Figure 8 shows a two-gimballed directional gyroscope as installed in U.S. Navy aircraft in World War II. The compass card showing degrees is affixed to the vertical gimbal and can be viewed through a small window next to which a lubber's line has been marked. The rotor is powered by jets of air. The caging knob releases the vertical gimbal to rotate, allowing the device to be reset with the rotor's axis horizontal and aligned with north. The use of a magnetic compass to align the rotor, however, may introduce compass errors of variation and deviation, requiring the pilot to understand the limitations of both instruments. A magnetic sensor known as a flux valve, found in later aircraft, transformed the directional gyroscope into a gyrocompass. The most common marine gyrocompasses have featured not two but three gimbals. When the ship's motion introduces a tilt, marine gyrocompasses compensate by the action of a pendulous mass that applies a torque to the rotor, causing the spinning axis to return to the north–south meridian. In fact, the tilt is necessary to the determination of the meridian by allowing the pendulous mass to cause

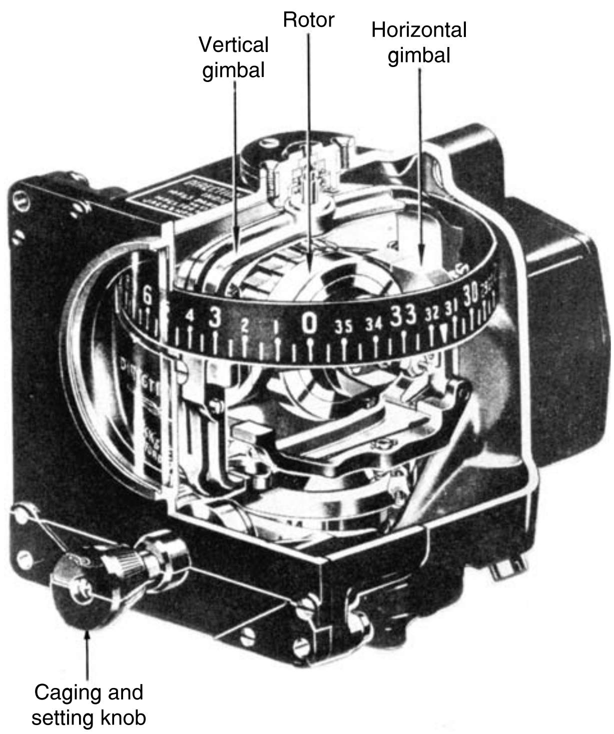


Figure 8. Gyroscopic direction finder used in U.S. Navy aircraft during World War II.
[*Source: U.S. Navy, Air Navigation Part Four: Navigation Instruments, Flight Preparation Training Series, McGraw-Hill, New York, 1944, p. 103.*]

the rotor to describe a diminishing circle about the pole, finally settling into the polar alignment. With no tilting, the gyrocompass ceases to work, a condition that occurs under some circumstances, precluding the marine gyrocompass from being applied in aircraft.

Although the Cold War arms race after World War II led to the development of inertial navigation, German manufacture of the V-2 rocket under the direction of Wernher von Braun during the war involved a proto-inertial system, a two-gimballed gyro with an integrator to determine speed. Inertial guidance combines a gyrocompass with accelerometers installed along orthogonal axes, devices that record all accelerations of the vehicle in which inertial guidance has been installed. With this system, if the initial position of the vehicle is known, then the vehicle's position at any moment is known because integrators record all directions and accelerations and calculate speeds and distance run. Inertial guidance devices can subtract accelerations due to gravity or other motions of the vehicle. Because inertial guidance does not depend on an outside reference, it is the ultimate dead reckoning system, ideal for the nuclear submarines for which they were invented and for ballistic missiles. Their self-contained nature makes them resistant to electronic countermeasures. Inertial systems were first installed in commercial aircraft during the 1960s. The expense of manufacturing inertial guidance mechanisms (and their necessary management by computer) has limited their application largely to military and some commercial purposes. Inertial systems accumulate errors, so their use at sea (except for submarines) has been as an adjunct to other navigational methods, unlike aircraft applications. Only the development of the global positioning system (GPS) at the end of the century promised to render all previous navigational technologies obsolete. Nevertheless, a range of technologies, some dating to the beginning of the century, remain in use in a variety of commercial and leisure applications.

See also **Global Positioning System (GPS); Radionavigation.**

ROBERT D. HICKS

Further Reading

Burger, W. and Corbet, A.G. *Marine Gyrocompasses and Automatic Pilots*, vol. 1. Pergamon Press, Oxford, 1963.

Defense Mapping Agency Hydrographic/Topographic Center. *The American Practical Navigator*. Pub. No. 9, DMAHTC, Bethesda, MD, 1995.

May, W.E. *A History of Marine Navigation*. W.W. Norton, New York, 1973.

Richardson, K.I.T. *The Gyroscope Applied*. Philosophical Library, New York, 1954.

Tetley, L. and Calcutt, D. *Electronic Navigation Systems*, 3rd edn. Butterworth-Heinemann, London, 2001.

Williams, J.E.D. *From Sails to Satellites: The Origin and Development of Navigational Science*. Oxford University Press, Oxford, 1994.

H

Hall Effect Devices

The "Hall effect," discovered in 1879 by American physicist Edwin H. Hall, is the electrical potential produced when a magnetic field is perpendicular to a conductor or semiconductor that is carrying current. This potential is a product of the buildup of charges in that conductor. The magnetic field makes a transverse force on the charge carriers, resulting in the charge being moved to one of the sides of the conductor. Between the sides of the conductor, measurable voltage is yielded from the interaction and balancing of the polarized charge and the magnetic influence.

Hall effect devices are commonly used as magnetic field sensors, or alternatively if a known magnetic field is applied, the sensor can be used to measure the current in a conductor, without actually plugging into it ("contactless potentiometers"). Hall sensors can also be used as magnetically controlled switches (see below), and as a contactless method of detecting rotation and position, sensing ferrous objects.

The first device employing the Hall effect was created in 1926 by Dr. Palmer H. Craig. The electromagnetic detector and amplifier were constructed of bismuth plates or films stacked together and wrapped with copper coils or wires. The device was able to supply radio reception without the use of batteries and vacuum tubes.

Hall effect switches combine Hall elements and a switching circuit. They are electronic switches operated by placing a magnet close to them. No direct physical contact is required, just close proximity to a magnet. The device can replace mechanical switches in various robotic and automotive applications. The use of a high-speed revolution counter can make a Hall effect switch operate at a high frequency. The advantages of Hall effect devices include their greater reliability compared to mechanical contact switches, the fact that, unlike optical sensors, they can function even when dirty (since they are contactless), and, because they are totally enclosed, they can operate in dirty environments, such as salt- and fuel-filled environments in automotive applications.

Some Hall effect devices have a binary output (on or off); others have a linear output. Devices with a linear output give voltage output proportional to the strength of the magnetic field and can provide an estimation of the distance from the magnet.

Hall effect devices have numerous electronic functions. They are used commonly as magnetic sensors such as a simple open or closed sensor. For example, in a seat belt sensor, the buckle, made of a ferrous material, interrupts the magnetic field between a magnet and the Hall effect device. When the field is interrupted, the device output switches on and when the buckle is removed the device switches off. This information is sent to the controller. Other automobile Hall sensors are electric windows and control systems for doors. Modern computerized engine control systems also use Hall effect sensors to sense crankshaft and camshaft speed and position. Again, the sensor detects a change in voltage when the magnetic field between a magnet and the Hall effect device is interrupted. A shutter rotating with the engine shaft passes through this opening, and with the change in voltage each time the shutter passes through the magnetic field, a rotation is counted.

The output Hall voltage is very small, and is usually amplified with several transistor circuits.

Chips based on the Hall effect became inexpensive and widely available in the 1970s, and eliminated the need for operators of the devices to design and manufacture their own magnetic circuits. In theory, Hall effect devices should have zero output voltage when no magnetic field is present, but in practice an ''offset'' voltage occurs due to material and mechanical defects. The offset voltage can be calibrated for, but may drift with temperature variations or mechanical stresses, masking low signals. Programmable linear Hall effect devices with active error correction provide better accuracy, linearity, and reliability with low thermal drift of the offset voltage.

The quantized Hall effect finds that the Hall resistance at low temperatures of a two-dimensional system (a thin sheet, such as in semiconductor transistors) exists as steps, rather than varying smoothly with voltage. The Hall resistance is said to be quantized. The quantum Hall effect, discovered by Klaus von Klitzing in 1980 is now the basis of very precise resistance measurements, needed to calibrate sensitive electronic test equipment. Quantum Hall devices produce very precise values of resistance of the order of 10 parts per billion.

See also **Computer Science; Electronics; Quantum Electronic Devices; Radio Receivers, Crystal Detectors and Receivers**

TIMOTHY S. BROWN

Further Reading

Kamimura, H. and Aoki, H. *The Physics of Interacting Electrons in Disordered Systems*. Clarendon Press, Oxford, 1989.

Lury, S., Xu, J. and Zaslavsky, A. *Further Trends in Microelectronics*. Institute of Electrical and Electronics Engineers, Wiley/Interscience, New York, 2002.

Putley, E.H. *The Hall Effect and Related Phenomena*. Butterworths, London, 1960.

Shrivastava, K.N. *Introduction to Quantum Hall Effect*. Nova Science Publishers, Hauppauge, NY, 2002.

Weisbuch, C. *Quantum Semiconductor Structures: Fundamentals and Applications*. Academic Press, Boston, 1991.

Health

The relationship between technology and health in the twentieth century has been multifaceted and complex. However, some general trends are observable. First, a variety of imaging techniques were pioneered during the course of the century, which permitted the *visual* representation of the living organism, including the ability to ''see'' the presence or absence of disease. Second, technologies were often ''disease specific'' in that they were designed with the intent of detecting or, in the case of vaccines, preventing particular diseases. Third, although most biomedical technologies have been beneficial, a few have come to be regarded as health threatening. Finally, the ''information technologies'' associated with the coming of the computer proved to be invaluable in recording and storing information about mortality and morbidity, which provided the foundation for public health policies. These trends are illustrated by examining some of the major health threats and biotechnological developments of the twentieth century. Although the focus will be on technology in the context of industrially developed Western societies, it is impossible to exclude other parts of the world when discussing these issues.

Perhaps the most famous twentieth century technology actually predated the emergence of the century by a little more than four years—the x-ray that was discovered by the German physicist Wilhelm Roentgen late in 1895. This was only the first in a series of ''visual technologies'' that permitted the examination of the inner workings of the human form. Two prominent examples from later in the century were computerized axial tomography (CAT) scan, developed by the British engineer Godfrey Newbold Hounsfield in the late 1960s and early 1970s, and the later technique of magnetic resonance imaging (MRI). Both techniques produced three-dimensional images of a human body on a screen. The 1970s also witnessed the introduction of positron emission tomography (PET) scanning. This technology used a radioactive tracer and permitted the observation of organ function (such as local blood flow) rather than focusing solely on organ structure as did the CAT and MRI scans. Although these technologies have become prominent symbols of the role of technology in understanding health and disease, they are technologies primarily available in the industrially developed countries of the world. According to one study by Tubiani in 1997, five sixths of the world's radiological equipment resides in countries that comprise one sixth of the world's population.

The x-ray soon proved its usefulness in screening for tuberculosis, one of the major health threats for much of the twentieth century. By the second decade of the century in the U.S., the newly formed National Association for the Study and Prevention of Tuberculosis urged that family and associates of a tubercular patient receive a precautionary chest x-ray. By the time of World War II, radiologists on

both sides of the Atlantic anticipated the prospect of "mass radiology of the chest." In the U.S., this was required of all military recruits after January 1941.

Although the primary function of x-ray examinations remained screening for tuberculosis, other diseases could be detected as well. As the century wore on, lung cancer became increasingly prominent. Technology played a role in the emergence of this disease in the sense that the mass-production of cigarettes was a distinctively twentieth century technology. Cigarettes became readily available on a wide scale and contributed significantly to lung cancer as a serious health threat. Even though there has been a reduction in the amount of cigarette smoking in Western societies since the health risks became known in the 1950s and 1960s, it is still a widespread practice in much of the developing world.

In addition to their obvious uses for diagnostic purposes, technologies have also proved to be useful therapeutically. Soon after Pierre and Marie Curie discovered the radioactive substance radium in 1898, it was put to medical use. This set the stage for the eventual development of radiation therapy as a standard treatment for diseases, most notably certain types of cancer. In 1917, the physicist Albert Einstein showed that under certain circumstances molecules could be stimulated to produce directed electromagnetic radiation. This insight laid the foundation for what later became the technology of the laser (that is, light amplification by stimulated emission of radiation). Eventually laser technology proved useful for restoring detached retinas and clearing blocked coronary arteries. Although not without risks, these technologies were instrumental in restoring individuals to health.

Whereas x-rays and lasers were dependent on developments in physics, the technologies that relied on chemistry were also prominent in contributing to health in the twentieth century. Like the x-ray, innovations in the last decades of the nineteenth century provided a context for later developments. In the 1880s, the French scientist Louis Pasteur and his German counterpart Robert Koch ushered in the science of bacteriology, which held that microorganisms were the cause of infectious disease. These microorganisms could be observed under a microscope in a laboratory setting and, after being isolated in a pure culture, could produce the disease anew in laboratory animals.

As can be deduced etymologically, Pasteur's work led to the process of pasteurization, which helped to preserve liquids like milk and wine and thereby contributed greatly to the public health. Because milk and milk products were a source of such diseases as tuberculosis, brucellosis, typhoid, diphtheria, salmonellosis, and streptococcal infections, the technologies of heat processing of food proved to be very important in contributing to a decline in milk-borne diseases throughout the twentieth century.

Among Koch's discoveries was the cholera bacillus organism that caused cholera. Because this bacillus was water-borne, Koch's work proved to be a spur to ensure that, for the sake of public health, communities receive clean water supplies. However, the association of a clean water supply with improved public health actually predated Koch's work—it had already been developed by physicians and sanitary reformers who had been politically active in Europe and the U.S. from the second quarter of the nineteenth century onward. In the twentieth century, the efforts to ensure a clean and health-sustaining water supply have increasingly relied on chemical substances like chlorine and fluoride, which have been introduced into water used for swimming and drinking. Although these developments produced major improvements in public health (e.g., the reduction in dental decay), they have also caused concern because of fears that chlorinated disinfection byproducts might prove to be carcinogenic.

In addition to practical developments like providing a scientific rationale for a clean water supply, Koch is generally credited with giving bacteriology a theoretical foundation. Although some remained skeptical of Koch's claim that infectious diseases were caused by specific microorganisms, his work generated many followers on both sides of the Atlantic. By the twentieth century, the bacteriological laboratory became a symbol of a new "scientific" public health that targeted specific "at risk" populations because of their exposure to specific germs. In the decades leading up to World War II, the search for bacteria through the techniques that Koch had pioneered was extended throughout the globe. This period coincided with a time of European colonization of much of the rest of the world, and the extension of Koch's techniques by European medical overseers was part and parcel of this broader movement.

One early use of the diagnostic laboratory from the early twentieth century onward was to screen for syphilis. In 1901, Jules Bordet and Octave Gengou developed a method of using blood serum to determine the presence of microorganisms, and five years later August von Wasserman and his

associates developed a blood serum test that detected the presence of syphilis. Within a very few years, local boards of health had added the Wasserman test to their list of standard procedures. In the U.S., many states required that couples undergo this test prior to marriage.

With the emergence of the modern pharmaceutical industry, the twentieth century witnessed the transformation of diabetes from a near fatal disease to a medically manageable condition. In the 1920s, Frederick Banting and Charles Best at the University of Toronto pioneered a method to develop insulin in a form that could be injected into the diabetic patient. When this fundamental scientific breakthrough was developed commercially by the Eli Lilly Company of Indianapolis, Indiana, it became possible for many diabetic individuals to lead comparatively normal lives. Because of the drug-induced potential to manage chronic conditions like diabetes, there were increasing calls for individuals to be screened for this condition on a regular basis.

By the end of World War II, major breakthroughs in industrially produced antibiotic drugs were occurring. Among the most prominent were the development of the drug streptomycin to treat tuberculosis and the multiple uses for the drug penicillin (e.g., to treat syphilis). Along with these success stories, there were also major health threats posed by the new and powerful drugs. In the latter category, probably the most famous example was the drug thalidomide. Produced commercially in Germany, it was widely prescribed in Europe in the late 1950s and early 1960s for women suffering from morning sickness. By 1961, pregnant women in 48 countries had taken the drug, and this resulted in the birth of over 8000 infants with deformities. This led to a ban on the drug in the U.S. and eventually to state testing and oversight of drugs sold by the pharmaceutical industry in most Western societies.

In the second half of the twentieth century, identification of chronic conditions such as hypertension and heart disease became more prominent. The technologies associated with detecting these health problems had been developed in the last years of the nineteenth century. In 1896, the Italian physician Scipione Riva-Rocci developed a method of measuring a patient's blood pressure by restricting blood flow in the arm through circular compression. Although there were later technical improvements, this approach to blood pressure measurement became the standard throughout Western medicine. In 1903, the Dutch physiologist Willem Einthoven developed the electrocardiograph, a technology that permitted the electrical activity of the heart to be monitored in an effort to detect cardiac disease. After the 1940s, routine testing for high blood pressure became more widespread as the concept of multiphasic screening (the attempt to locate disease in seemingly well people by giving them a wide variety of tests) became common. By the last quarter of the century, a whole host of cardiovascular agents were developed in an attempt to make hypertension and other heart-related illnesses into manageable health problems.

This shift to a concern with heart disease vividly illustrates the key epidemiological transition of the twentieth century; that is, the shift from infectious to chronic diseases as the leading causes of mortality and morbidity. Chronic conditions take years to develop and often result from lifestyle choices regarding diet and exercise rather than exposure to an infecting agent. Both diet and exercise have produced their attendant technologies. In response to these changing health realities and consumer demand, food producers have created foods that are lower in fat and refined sugar, higher in fiber, and "fortified" with vitamins and minerals. Similarly, the concern with exercise has spawned the creation of health clubs with a variety of machines designed to produce a healthy body in addition to squash and tennis courts, swimming pools, exercise bicycles, and running/walking tracks.

The postwar era also saw the emergence of vaccines, which led to the eradication of many diseases. In 1960 American researcher John Enders attempted to develop a measles vaccine, which was licensed in 1963 and is credited with saving a large number of lives. Similar vaccines were developed for mumps and rubella, and it is now possible in most developed countries to receive a triple vaccine that immunizes the individual against all three conditions.

Polio is one of the most famous examples of disease eradication by means of vaccination. In the 1950s, American researchers Jonas Salk and Albert Sabin both developed polio vaccines. The Salk vaccine used an inactivated (killed) polio virus whereas Sabin used a live-attenuated virus that could be administered orally. With support from the National Foundation for Infantile Paralysis, industrial production of the Salk vaccine began in the U.S. immediately after its effectiveness had been established in a clinical trial conducted in 1955. In European countries, Salk's vaccine was either imported from the U.S. or developed in government laboratories. By the 1970s, however,

Salk's vaccine had been supplanted by Sabin's. United States pharmaceutical companies ceased producing the Salk vaccine, and the World Health Organization began to issue the Sabin vaccine to developing countries.

The development of vaccines was also of vital importance in the eradication of smallpox. As C.C. Booth has observed, the discovery of a method for producing a freeze-dried vaccine was an indispensable aspect of this endeavor. Following a worldwide vaccination campaign, smallpox was confined to India, Ethiopia, Somalia, and Botswana by 1973. By 1977, the disease had been removed from Ethiopia and Somalia, and the official eradication of the disease from the world was declared at a 1980 meeting of the World Health Organization. As the chairman of the commission overseeing the eradication campaign observed, this was an event that was "unique in the history of mankind". Despite this optimism, at the dawn of the twenty-first century concerns still linger that the few existing samples of the smallpox virus might be used for malevolent purposes should they fall into the wrong hands.

In the last two decades of the twentieth century, the new life-threatening disease of AIDS (acquired immune deficiency syndrome) emerged on the world stage. Although initially diagnosed in homosexual men and intravenous drug users in the U.S., it spread throughout the world with approximately 70 percent of the cases in sub-Saharan Africa; also, at the beginning of the twenty-first century, most infections result from heterosexual transmission. Based on whether one lives in the developed or developing world, AIDS is a different disease. In the industrially developed world, it has become a chronic condition that, although still life threatening, can be treated with drugs that interfere with viral replication. By contrast, in the developing world, AIDS remained a major killer with many countries lacking adequate financial resources to purchase the drugs that can prolong the life of infected individuals. To help ensure that all the world's people (and not just those in the developed parts) have access to vaccines, the International AIDS Vaccine Initiative (IAVI) was established in 1996. By setting up a Vaccine Purchase Fund, the IAVI hoped to establish that there will be a guaranteed market in the developing world and thereby create incentives for private sector vaccine companies to enter this market. However, all these developments are predicated on being able to develop an effective AIDS vaccine, and at the time of writing none of these efforts has been successful.

Although most biomedical technologies have been developed with the express intent of improving health, many technologies have also been perceived as health risks. From early on, the potential dangers of x-ray exposure were reported. Similarly, the technology of electronic fetal monitoring has had a varied history. Pioneered in the 1950s and 1960s by researchers in the U.S., Germany, and South America, the technology was heralded as a way to detect fetal distress during delivery. Even though this claim has been the subject of much debate within the medical community, the widespread use of the practice has had the effect of increasing the number of deliveries by caesarean section, which, like any surgical intervention, constitutes a health risk for the mother. Other technologies that have been alleged to be health risks (often in a court of law) include exposure to electromagnetic fields and silicone breast implants. Although many in the scientific community have questioned the empirical basis for these claims, the historically important fact is that the legal arena has become a prominent forum in which the health risks of technology are actively debated.

In addition to considering how technologies help to determine whether a particular individual is in a state of health or disease—as revealed by an x-ray image or blood pressure test—the more abstract concepts of "health" are dependent on technology as well. In particular, "information technologies" have been developed to retrieve and preserve data in individual health records and in aggregate data about populations. Of course, the collecting of aggregate data predates the twentieth century, but computer-based technologies have permitted this practice to be carried out on a much wider scale. By the 1990s there were calls on both sides of the Atlantic for health care providers to practice "evidence-based medicine," that is, practice based on empirical evidence derived from studying populations. In this respect, technology has gone beyond merely revealing the presence or absence of disease and has become ineluctably interwoven with how we conceive of health and disease in contemporary Western society.

See also **Antibacterial Chemotherapy; Biotechnology; Cardiovascular Disease, Diagnostic Screening; Genetic Screening and Testing; Immunization Technology; Medicine; Nuclear Magnetic Resonance (NMR, MRI); Positron Emission Tomography (PET); Tomography in Medicine**

J. Rosser Matthews

Further Reading

Annas, G.J. and Elias, S. Thalidomide and the *Titanic*: reconstructing the technological tragedies of the twentieth century. *Am. J. Pub. Health*, 89, 1, 98–101, 1999. This article is written as a policy piece and argues that, suitably regulated, the drug thalidomide could be reintroduced into the American market; however, it also provides the historical background to the thalidomide tragedy.

Banta, H.D. and Thacker, S.B. Historical controversy in health technology assessment: the case of electronic fetal monitoring. *Obst. Gynecol. Surv*., 56, 11, 707–719, 2001. This article historically recounts how electronic fetal monitoring was rapidly diffused within obstetrical practice and discusses the debates that ensued when the authors published a 1979 study, which indicated that there was insufficient evidence to recommend its use as a standard procedure.

Berkley, S. HIV vaccine development for the world: an idea whose time has come? *AIDS Res. Hum. Retroviruses*, 14, s. 3, 191–196, 1998. The author advocates that an AIDS vaccine be developed and discusses possible financing arrangement like the International AIDS Vaccine Initiative.

Blume, S. and Geesink, I. A brief history of polio vaccines. *Science*, 288, 5471, 1593–1594, 2000. This article traces how the Sabin vaccine against polio eventually supplanted the Salk vaccine; the authors use this story as a case study of the process of technological ''lock in,'' a view of technological change that they attribute to evolutionary economists.

Booth, C. Changing concepts of health and disease during the 20th century. *J. R. Coll. Phys. London*, 34, 1, 38–46, 2000. As the title (correctly) indicates, this article provides a nice overview of changing concepts of health and disease-including discussions of the role that biomedical technologies have played in this process.

Brandt, A.M. The cigarette, risk, and American culture. *Daedalus*, 119, 4, 155–176, 1990. Among other things, this article discusses how technological forces shaped ''the rise of the cigarette'' in twentieth century American culture, with its attendant health risks.

Caplan, A.L. The concepts of health, illness, and disease, in *Companion Encyclopedia of the History of Medicine*, Bynum, W.F. and Porter, R., Eds. Routledge, London, 1993. This article provides a philosophical overview of changing conceptions of health and disease, including a discussion of how technology can ''objectify'' the patient.

Daly, J. and McDonald, I. Cardiac disease construction on the Borderland. *Soc. Sci. Med*., 44, 7, 1997, 1043–1049. Adopting a social constructionist perspective, the authors discuss the role of technology in defining whether an individual suffers from heart disease; the authors survey developments throughout the twentieth century.

Desrosières, A. *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press, Cambridge, MA, 1998. Original French edition, 1993 As a work that attempts to synthesize a vast amount of the scholarly literature on the history of statistics, this book would be a good source for bibliographic leads on the history of medical and public health statistics in the twentieth century.

Higgs, E. The statistical big bang of 1911: ideology, technological innovation, and the production of medical statistics. *Soc. Hist. Med*., 9, 3, 409–426, 1996. By discussing what he terms ''the nuts and bolts of statistical creation,'' the author illuminates the prehistory that set the stage for the computer revolution in data collection later in the century; the geographical focus in on early twentieth century Britain.

Kennedy, R.G. Electronic fetal heart rate monitoring: retrospective reflections on a twentieth-century technology. *J. R. Soc. Med*., 91, 244–250, 1998. The author discusses the history of how electronic fetal monitoring was introduced into British obstetric practice with a focus on why it disseminated so rapidly.

Marks, H.M. Medical technologies: social contexts and consequences, in *Companion Encyclopedia of the History of Medicine*, Bynum, W.F. and Porter, R., Eds. Routledge, London, 1993. A good overview with a focus on developments from 1850 to the present.

Porter, D. *Health, Civilization and the State: A History of Public Health from Ancient to Modern Times*. Routledge, London 1999. Although the title implies an encyclopedia overview, the focus is on developments in public health in the nineteenth and twentieth centuries. There is a discussion of ''healthy'' lifestyles and attendant developments; for example, bodybuilding.

Porter, R. *The Greatest Benefit to Mankind: A Medical History of Humanity*. W.W. Norton, New York, 1999. A panoramic overview of the history of Western medicine, which includes discussions of developments in health and technology in the twentieth century.

Reiser, S.J. *Medicine and the Reign of Technology*. Cambridge University Press, Cambridge, 1978. This book surveys technological developments from early modern times (sixteenth to the eighteenth centuries) to contemporary concerns.

Reiser, S.J. The emergence of the concept of screening for disease. *The Milbank Memorial Fund Quarterly*, 56, 4, 403–425, 1978. This article provides a nice overview of how the concept of ''screening'' populations to determine their health status emerged in twentieth century American society.

Reiser, S.J. The science of diagnosis: diagnostic technology, in *Companion Encyclopedia of the History of Medicine*, Bynum, W.F. and Porter, P, Eds. Routledge, London, 1993. Although covering some of the same material as his book, Reiser brings the story up to more recent developments.

Sepkowitz, K.A. AIDS—the first 20 years. *New Engl. J. Med*., 344, 23, 1764–1772, 2001. This article attempts to review the main features of the response to AIDS in the last 20 years.

Tubiana, M. From Bertha Roentgen's hand to current medical imaging: one century of radiological progress. *Eur. Radiol*., 7, 9 1507–1513, 1997. This article provides a historical overview of the development of radiology and advocates that the technology be extended to the developing world.

Welch, R.W. and Mitchell, P.C. Food processing: a century of change. *Br. Med. Bull*., 56, 1, 1–17, 2000. As the title states, this article surveys developments in food processing technology during the course of the twentieth century—drying, heat processing, freezing and chilling, controlled and modified atmospheres, radiation, and mechanization/standardization. Also, the health implications of changes in food processing are discussed.

Hearts, Artificial

Heart disease is a prominent cause of death in affluent nations. In the U.S., it is the number one killer and accounts for more than 39 percent of all deaths. Treatment for heart failure includes medical therapy, surgical intervention, cardiac transplantation, and the use of cardiac devices. Cardiac devices such as heart valves, pacemakers, defibrillators, and intra-aortic balloon pumps can assist the heart, and an artificial heart can replace it.

There are two groups of artificial hearts: total artificial hearts and partial artificial hearts. The total artificial heart (TAH) replaces the failing ventricles (pumping chambers) of a damaged human heart, which is excised or removed from the body. A partial artificial heart or ventricular assist device (VAD) attaches to a failing heart (which remains in the body) and serves to assist in the pumping function of the heart. The main components of both TAHs and VADs are the energy source, drive-control unit, actuator, and pump. Historically, the actuation of artificial hearts has included pneumatic, hydraulic, electrical (electrohydraulic, electromechanical, electromagnet) and nuclear (thermoenergy) power. Nuclear artificial heart development was halted due to individual and societal concerns relating to the plutonium radioisotope power source. The first artificial hearts were constructed out of polyvinylchloride, which tested well for endurance, and later silastic and polyurethane materials. Improved biomaterials, particularly compatible blood surface materials, continue to be developed. Over time, both total and partial artificial hearts have changed from large devices situated outside the body (paracorporeal and extracorporeal) to smaller devices intended to be placed inside the body (intracorporeal and fully implantable).

Willem Kolff and Michael DeBakey are among the pioneers in the field of artificial heart development. Kolff invented the artificial kidney machine in Nazi-occupied Holland in 1943. After the war, Kolff led an artificial organ research program at the Cleveland Clinic in Cleveland, Ohio that encouraged research into artificial kidneys, eyes, ears, arms, lungs, and hearts. In 1957, Kolff and Dr. Tetsuzo Akutsu successfully implanted a crude air-driven artificial heart, composed of two polyvinylchloride pumps, in a dog and maintained circulation for ninety minutes. The Cleveland Clinic group experimented with hydraulically-activated, electromechanical and pneumatic-driven TAHs, and later various types of VADs. At the Baylor College of Medicine in Houston, Texas, DeBakey and his research team developed a different type of TAH. Their device consisted of two pneumatically-driven silastic sac-type pumps implanted for biventricular bypass. The Baylor College of Medicine group also expanded their research into VADs, heart valves and vascular prostheses.

DeBakey was instrumental in lobbying for the formation of the U.S. Artificial Heart program, established in 1964, at the National Heart Institute of the National Institutes of Health (NIH) in Bethesda, Maryland. The NIH became an important funding source and catalyst for artificial heart research. Under the initial direction of Frank Hastings, the NIH artificial heart program supported investigator research grants and industry contracts for the development of assist and replacement devices totaling more than $400 million. Outside of the U.S., artificial heart research programs were also established in Germany, the former Soviet Union, Japan, and other countries whose research teams experimented with various TAH and VAD prototypes.

By the late 1960s, there were four institutions in the U.S. working on developing TAHs for clinical application: Willem Kolff's team, now at the University of Utah in Salt Lake City, Utah; the Baylor College of Medicine group headed by Michael DeBakey; William Pierce and his research team at Pennsylvania State University; and the Cleveland Clinic under the direction of Yukihiko Nosé. These researchers contributed incremental improvements on device designs, materials, performance, durability and outcome measures through extensive animal experiments.

The device used in the first clinical case was the pneumatic Liotta TAH, developed by Domingo Liotta and tested in the Baylor College surgical laboratory on calves. At the Texas Heart Institute, Denton Cooley implanted the device as a bridge-to-transplantation in 47-year-old Haskell Karp who could not be weaned from cardiopulmonary bypass after heart surgery. The Liotta TAH sustained Karp for 64 hours, at which time the patient underwent cardiac transplantation. In 1981, Cooley implanted the pneumatic Akutsu III TAH, developed by Tetsuzo Akutsu, which provided a patient with 39 hours of support as a bridge-to-transplantation. Both device implant cases were technical successes, but both patients died shortly after the transplant surgery.

In 1982, William DeVries implanted the pneumatic Jarvik-7 TAH, developed by Robert Jarvik and fellow researchers at the University of Utah, in Barney Clark. Ineligible for cardiac transplanta-

tion, Clark consented to the TAH as a permanent implant. During his 112 days with the TAH, Clark experienced numerous complications including device problems, nasal bleeding and neurological incidents. In 1984 and 1985, three more patients in the U.S. and one in Sweden received permanent Jarvik-7 TAH implants, surviving from 10 to 620 days. In response to severe patient complications, the U.S. Food and Drug Administration suspended Jarvik-7 TAH production and its use as a permanent device in 1991. The Jarvik-7 TAH (renamed Symbion, later CardioWest) continued however to be implanted in patients as bridge-to-transplantation with moderate success rates.

Knowledge gained in TAH research contributed to VAD advancements. VADs experienced success as short-term mechanical support systems, bridge-to-transplantation, and possibly permanent cardiac devices. DeBakey's research group developed a pneumatic paracorporeal (outside the body) left ventricular assist device (LVAD) for short-term ventricular assistance. In 1967, the first clinical case was reported in which a patient, unable to be weaned from cardiopulmonary bypass, temporarily utilized the LVAD and then had it removed upon cardiac recovery. By the mid-1970s, clinical trials were underway at the Texas Heart Institute and Boston Children's Hospital to evaluate the intracorporeal pneumatic VAD—the TECO Model VII—as both a temporary recovery measure and bridge-to-transplantation. In 1984, the electrically powered left ventricle assist system—the Novacor LVAS—emerged as both the first implantable system and successful bridge-to-transplant clinical case (Figure 1). In that same year, the paracorporeal, pneumatic Thoratec VAD also reported a successful bridge-to-transplantation case. Further refinements in the 1990s brought forth wearable ventricular assist systems, which allowed patients to resume basic activities. Clinical trials also began to evaluate VADs for permanent use in patients ineligible for cardiac transplantation.

In 2001, clinical trials with the electromechanical AbioCor TAH as a permanent device began in the U.S. The AbioCor TAH is the first fully implantable device; its components of the replacement heart (hydraulically-driven pump), controller and rechargeable internal battery are implanted in the patient's chest and abdomen. The AbioCor TAH utilizes a transcutaneous energy transmission (TET) device as a wireless energy transfer system without need for percutaneous lines risking infection.

Artificial heart research encompasses the development of total artificial hearts (TAHs) (see Figure 2) and partial artificial hearts (VADs). In both cases, these devices became more sophisticated, smaller in size, and fully implantable. Both devices were explored as temporary and permanent therapies. Whereas the TAH remains an experimental device, several ventricular assist systems are now

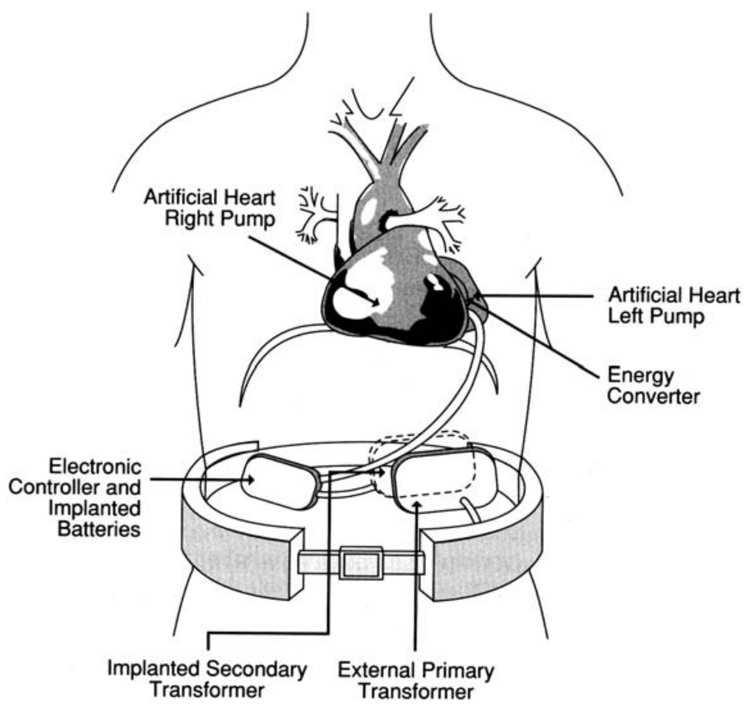


Figure 1. The Novacor fully implantable ventricular assist device.
[*Used with permission from The Artificial Heart: Prototypes, Policies, and Patients, by the National Academy of Sciences, courtesy of the National Academy Press, Washington D.C., 1991.*]

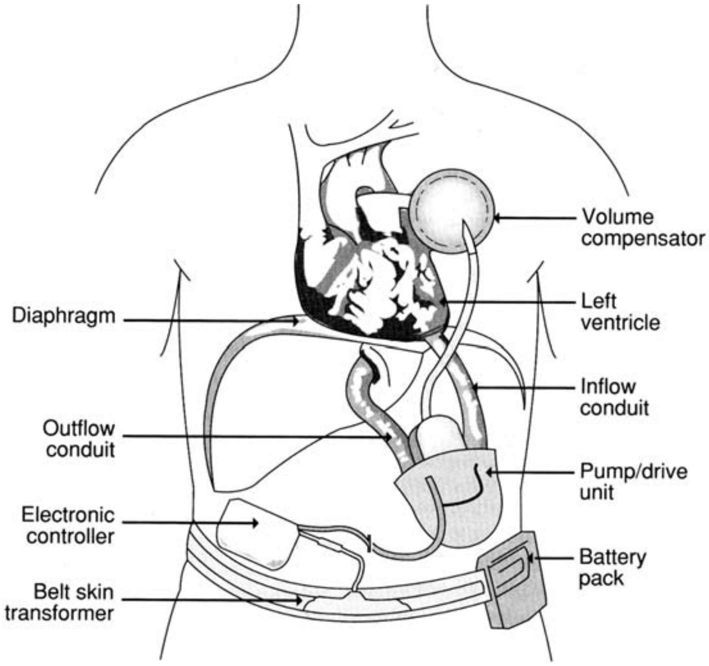


Figure 2. A fully implantable total artificial heart. [*Used with permission from The Artificial Heart: Prototypes, Policies, and Patients, by the National Academy of Sciences, courtesy of the National Academy Press, Washington D.C., 1991.*]

FDA-approved for clinical use. Not without debate, artificial heart technologies have raised contentious issues of access, quality of life, cost-benefit and reimbursement. The broader context is the role that artificial heart technologies may play towards continuing or challenging the technological imperative in medicine.

See also **Biotechnology; Cardiovascular Surgery; Implants, Heart Pacemakers, and Heart Valves**

SHELLEY MCKELLAR

Further Reading

Cauwels, J.M. *The Body Shop: Bionic Revolutions in Medicine*. Mosby, St. Louis, MO, 1986.

Fox, R.C., Swazey, J.P. and Watkins, J.C. *Spare Parts: Organ Replacement in American Society*. Oxford University Press, New York, 1992.

Fox, R.C. and Swazey, J.P. *The Courage to Fail: A Social View of Organ Transplants and Dialysis*, 2nd edn. Transaction, New Brunsick, NJ, 2002.

Hogness, J.R., Van Antwerp, M. and National Heart Lung and Blood Institute, Eds. *The Artificial Heart: Prototypes, Policies, and Patients*. National Academy Press, Washington D.C., 1991.

Lubeck, D.P. and Bunker, J.P. *The Implications of Cost-Effectiveness Analysis of Medical Technology. Case Study Number 9: The Artificial Heart: Cost, Risks and Benefits. Background Paper Number 2: Case Studies of Medical Technologies*. U.S. Government Printing Office, Washington D.C., 1982.

Plough, A.L. *Borrowed Time: Artificial Organs and the Politics of Extending Lives*. Temple University Press, Philadelphia, 1986.

Preston, T.A. The artificial heart, in *Worse than the Disease: Pitfalls of Medical Progress*, Dutton, D.B., Ed. Cambridge University Press, New York, 1988, pp. 91–126.

Shaw, M.W. *After Barney Clark: Reflections on the Utah Artificial Heart Program*. University of Texas Press, Austin, 1984.

Helicopters

The term ‘‘helicopter’’ is derived from the ancient Greek ‘‘elikoeioas,’’ which means winding, and ‘‘pteron’’ meaning feather. Its first documented use came in the 1860s when Frenchman Ponton d’Amécourt tested miniature steam-powered models. His experiments, however, were built on a notion dating back two millennia.

The notion of vertical flying and hovering probably originated in the manufacture of toys consisting of leaves or feathers attached to a small stick and spun between the hands before being released. The earliest evidence of such toys is found in the form of Chinese tops dating back to 400 BC. Such toys gained in sophistication beginning in the eighteenth century, when Enlightenment experimenters tried to attach spring devices to help the contraption rise.

Several paper projects also existed, most notably Leonardo da Vinci’s 1483 sketch of an air screw pointed vertically. Though inspiring much speculation about its feasibility, more recent interpretations of the design suggest da Vinci, who may have

tested a small model of his design, was not seriously considering building such a machine.

By the nineteenth century, however, several aviation pioneers had further focused their efforts on the necessary elements to build rotating wings. The lack of an effective power plant made any machine impossible to build, but such pioneers as Englishmen George Cayley and W. H. Phillips, through their articles and model experiments, lay the groundwork for practical tests in the early twentieth century.

Most practical experiments were short-lived. Indeed, it became clear that the vertical rotor produced torque, shifting the rest of the machine around its axis. Neutralizing this torque effect while installing directional control became a central concern of experimenters. Solutions ranged from Thomas Edison's testing of rotors in the 1880s, to Italian Enrico Forlanini's dual counter-rotating propellers; yet all attempts confirmed the need for both a high-powered engine and a better understanding of aerodynamics.

In France, Louis Bréguet and Paul Cornu both built a machine capable of hovering, but their 1907 flights failed to control their respective machines' direction.

Several helicopter pioneers were eventually successful in devising solutions to the problem of directional control. Spaniard Juan de la Cierva, suggested a hybrid airplane/helicopter solution known as the autogiro. Using the principle known as autorotation, which capitalized on the phenomenon affecting fixed wings, Cierva transferred the principle to rotating surfaces. By adding wings and a propeller engine, and perfecting later models with collective pitch control (which allowed the pilot to determine lift, thrust and rotor rotation speed), Cierva launched a craze. The "flying windmill" as it became nicknamed was sold as a kind of model T of the air, though its speed was limited and it depended on proper wind conditions to lift-off. The death of Cierva in a plane accident in 1936 also signaled the decline of his "autogiro" formula.

That same year, the French Bréguet–Dorand machine set several endurance and hovering records, but still failed to demonstrate effective control capacity. Meanwhile the German Focke Fa-61 with two counter-rotating rotors took to the air and broke several of the French records. In the hands of female pilot Hanna Reitsch, it became a remarkable propaganda machine during 1938, and paved the way for the development of model Fa-223. But true German success came with Anton Flettner's model Fl-282 Kolibri, lighter than the Focke machine, and maneuverable to the point where it avoided simulated German fighter attacks, and became the first machine to reach a practical stage (and was ordered by the German navy).

The full breakthrough would come from the U.S., where Russian Igor Sikorsky had immigrated in 1919. The young inventor had been experimenting with helicopter formulas ever since returning from a trip to Paris in 1910, where he had learned of Bréguet's helicopter attempts. Through trial and error, Sikorsky began defining further the dynamics required to control lift-off. Finally in 1931, he applied for a basic helicopter patent that included a rotor with a small vertical propeller in the tail intended to counteract the torque effect created by the big rotor. The patent was granted in 1935 and Sikorsky's prototype, model VS-300, was rolled out in fall 1939. This seminal machine served as the basis for the design of the XR-4 model, which first flew in 1942 and was ordered into serial production for use in Burma in World War II, where the machines rescued downed pilots.

By the end of World War II several other helicopter manufacturers had entered the fray, including Bell helicopters, whose chief engineer Arthur Young helped design the highly successful Bell-47, which first appeared in 1949. By the time of the Korean War, helicopters began serving as medical evacuation machines, and soon after were used for commando drops. The performance of these machines, though steadily increasing, remained limited by the use of piston engines, which often experienced overheating. This changed with the introduction of turbine engines; that is, jet-power affixed to a propeller or a rotor. The French-built SE-3130 Alouette II was the first serial-built machine to use a jet engine, the turboshaft Artouste II.

Steady progress continued, yet helicopters did not evolve much with regard to the lift and control devices they use, which remained remarkably similar to those of Sikorsky's VS-300. A helicopter rotor includes two to five blades radiating symmetrically from a hub atop the machine. These function as both wings and propellers, which are driven by an engine. Because of a series of gears, the speed of the rotor blades is less than that of the engine. Unfortunately, this reduction on speed increases the phenomenon of rotational inertia, or torque, which pushes the helicopter in the direction opposite of that of rotation. This is usually countered through the installation of a smaller vertical rotor in the tail. Some models, however, incorporate a second rotor turning in the opposite direction, a solution adopted on several American and Russian models.

The lifting phenomenon functions in a manner similar to that of a plane's wings. However, to replace the rudders and ailerons, the blades (hinged to remain flexible) can rotate along their axis depending on which flight phase the helicopter is in. The rotor can modify the angle of the blades cyclically (each blade in turn as it completes a rotation), or simultaneously.

Piloting a helicopter relies on three primary controls. The "stick," similar to that in an airplane cockpit, is the cyclic pitch lever, which allows the aircraft to move forward, backward, or sideways; the collective pitch lever, on the left side of the pilot's seat, which controls the angle of descent and climb; and the rudder pedals.

Since the helicopter's flying depends entirely on the rotor, any engine shutdown can have catastrophic consequences. A safety feature that allows a rapid controlled descent is autorotation, where the gears are released, allowing the rotor to spin solely under the force of air, without engine input. This produces some lift, while safeguarding control functions.

The features of the modern-day helicopter have also limited its speed. Some combat machines, such as the Bell AH-1 Cobra gunship used in Vietnam, broke speed records, but their achievement remains an exception rather than the norm.

Soon after World War II, several airlines in the U.S. and Europe began applying for licenses to operate helicopter service, either from city to city or within a city, from downtown to the airport. Most of these scheduled services disappeared by the 1970s as a consequence of safety concerns (including several high-profile accidents in the U.S.), but also due to cost: the number of bookings for such flights failed to balance the high operational overhead that the maintenance of small heliports required. Instead, most helicopter lines rely on a combination of charter, sightseeing, and luxury services for high-end clientele to survive. Others specialize in specific hauling services, such as logging, heavy construction, and medical evacuation. The helicopter has also proven extremely useful to law enforcement in solving ground traffic problems and policing crime.

Though the overall shape of the helicopter has not changed much since Sikorsky's original patent, various innovations and materials have contributed to improving its performance. These range from the use of carbon composites, to placing the tail rotor within the axis of the tail to reduce noise. More recently, the use of directional outlets to blow air out of the tail (the NOTAR principle) has appeared as a way to replace the tail rotor.

Finally, newer combat models are displaying stealth capabilities. A recent innovation has involved the use of tilt rotor technology, whereby a vertical take-off machine tilts its engines and uses them in the manner of an airplane. The Bell Boeing V-22 Osprey is the latest incarnation of such an approach. Though touted as highly promising (especially in terms of speed gains and carrying capacity), it has thus far proven difficult to master operationally.

See also **Aircraft Design**

GUILLAUME DE SYON

Further Reading

Boulet, J. *History of the Helicopter: As Told by its Pioneers*. France-Empire, Paris, 1984.

Brooks, P.W. *Cierva Autogiros: The Development of Rotary-Wing Flight*. Smithsonian Press, Washington D.C., 1988.

Focke, H. German Thinking on Rotary Wing Development. *J. R. Aeronaut. Soc.*, May 1965.

Leishmann, J.G. *Principles of Helicopter Aerodynamics*. Cambridge University Press, Cambridge, 2000.

Liberatore, E.K. *Helicopters Before Helicopters*. Krieger, Malabar, FL, 1998.

Spenser, J.P. *Whirlybirds. A History of the U.S. Helicopter Pioneers*. University of Washington Press, Seattle, 1998.

Hematology

Hematology is the medical specialty that deals with conditions of the blood and blood-forming organs. Plentiful in supply and relatively easy to access, the blood is a tissue well suited to manipulation and investigation. Nineteenth century innovations in microscopy, such as the introduction of new staining techniques and phase-contrast methods, brought great advances in analysis of the blood. Such advances, in combination with the mass-production of the relatively easy to use hemocytometer and hemoglobinometer (used for measuring the size and number of blood cells and hemoglobin concentration, respectively), meant that the morphological and quantitative analysis of the blood became a fashionable part of practice for many early twentieth century physicians, especially those wishing to demonstrate familiarity with the latest methods in '"scientific medicine," reflecting and stimulating changes in medical practice and research more widely. It was during this time in North America and Europe that new institutional and intellectual ties between clinical medicine and basic science were forged, dramatically affecting the nature of clinical research. The study of the form and functions of the blood in health and

disease was a popular subject of research, attracting the attention of chemists and pathologists, research-oriented clinicians, and chemical physiologists. Similarly, as knowledge of the blood and its constituents increased, so routine diagnostic analysis of the blood became a central part of the work of the many new hospital laboratories springing up in this period across Europe and North America.

Between the mid-nineteenth and mid-twentieth centuries, research effort in hematology was particularly focused on the causes and treatment of the anemias. These were conditions commonly seen on the wards, which often provided great potential for analysis using the latest technologies. Sickle cell anemia among African–Americans, for instance, was formally identified by Chicago physician James Herrick in 1910, but as a rare disease found solely in members of an underprivileged minority population, it did not initially trigger much clinical or biological interest in the U.S., and still less in Europe, with its tiny black population. Nonetheless, by the 1930s, sickle cell anemia, along with thalassemia, another inherited hemoglobinopathy (first identified as a specific disease in 1925 by the Detroit pediatric physicians Thomas Cooley and Pearl Lee), was recognized as a genetically linked condition, and as such was fitted well in the growth of interest in genetic and molecular views of disease in the twentieth century. In 1945, chemist Linus Pauling and his new medically-trained graduate student Harvey Itano took up the problem of sickling in red blood cells as a tool through which to explore the chemical nature of hemoglobin, making sickle cell anemia the first disease to be fully described on a molecular level.

The link between iron deficiency and anemia, the root of one of the commonest of all nutritional deficiencies, was first made by Gustav von Bunge, Professor of Physiological Chemistry at the University of Basle in 1902. It was not until the 1930s, however, that British physicians Leslie Witts, D.T. Davies, and Helen MacKay each confirmed the role of nutritional iron in the replacement of iron contained in hemoglobin, and discussed the dietary lack or gastric malabsorption of this iron as leading to a specific disease process. In other anemia research, dietary treatment was also producing a great deal of research activity. In 1925 the Rochester physiologists George Whipple and Frieda Robscheit-Robbins published results showing that administration of beef liver to dogs with severe experimental anemia markedly increased the rate of blood regeneration, and they discussed the possible applications of this work to patients with pernicious anemia. Their work was complemented on the clinical side by George Minot and William Murphy, two Harvard medical school physicians, who announced in 1926 a dietary treatment for pernicious anemia through the administration of raw liver. As had been the case for insulin treatment for diabetes (discovered a few years earlier), these findings were celebrated as a wonderful example of the power of combined laboratory and bedside research, and in 1934, Whipple, Minot and Murphy were awarded the Nobel Prize in Medicine for their work. The anti-anemic factor in liver was isolated and identified as vitamin B12 in 1948 by two teams working independently, one British, led by E. Lester Smith, head of Glaxo's biochemistry department, and one American, led by Merck's director of biochemistry, Karl Folkers. The vitamin's chemical structure was worked out during the mid-1950s by Dorothy Hodgkin, the Oxford x-ray crystallographer, for which work she received the Nobel Prize in Chemistry in 1964.

The absence of some "intrinsic factor," (i.e., a substance required to absorb the anti-anemic factor), was first proposed by William Castle to be the underlying cause of pernicious anemia arising from his work at the Boston City Hospital during the 1930s (Castle also contributed significantly to studies of hypochromic anemia and iron deficiency ongoing there at this time also). Castle's idea stimulated enormous amounts of further hematological research aimed at the isolation and identification of this mysterious entity, and then, it was hoped, the development of a cure for pernicious anemia. Today, the disease remains incurable but treatable, and is now understood as an autoimmune disease. The 1930s also saw the introduction of a pioneering method of quantitative analysis of the blood, devised by the U.S. physician Maxwell Wintrobe, making use of his new invention, the "Wintrobe tube," designed to measure the volume of packed red cells and erythrocyte sedimentation rate. From this analysis, Wintrobe produced a classification of the anemias that endures to the present day.

If anemia defined the study and practice of hematology in the first half of the twentieth century, then advances in the understanding of blood malignancy (the leukemias and lymphomas) and the introduction of effective chemotherapies firmly established hematology as a formal specialty in the post-World War II years. The first professional hematological society, the International Society of Hematology, was founded in 1946, followed by the American Society of Hematology in 1954, and then several national societies across

Europe and Asia. New journals such as *Blood* (1946), *The British Journal of Haematology* (1955) and *Progress in Hematology* (1956) became crucial professional vehicles for the nascent specialty. The new oncology and pediatric oncology bodies and journals emerging during the 1960s and 1970s further extended the professional interactions of hematologists.

By the end of the twentieth century, a career in hematology required first training as either a physician or a pathologist, and involved some or all of the following: the management of blood products and derivatives; the administration of immunosuppressives, chemotherapies, anticoagulants and antithrombotic agents; and supportive care for a range of systemic diseases. On the research side, hematologists are actively engaged in research on several fronts including anemia, cancer and chemotherapy, blood typing and blood products, pain management studies, and stem cell therapy.

See also **Blood Transfusion and Blood Products; Cancer, Chemotherapies; Diagnostic Screening**

HELEN K. VALIER

Further Reading

Foster, W.D. *A Short History of Clinical Pathology*. London, 1961.

Jaffe, E.R., The evolution of the American and International Societies of Hematology: a brief history. *Am. J. Hematol.*, 39, 67–69, 1992.

Loudon, I. The history of pernicious anaemia from 1822 to the present day, in *Health and Disease: A Reader*, 1st edn. Milton Keynes, 1984.

Robb-Smith, A.H.T. Osler's influence on haematology. *Blood Cells*, 7, 513–533, 1981.

Wintrobe, M.M., Ed. *Blood, Pure and Eloquent: A Story of Discovery, People and of Ideas*. McGraw-Hill, New York, 1980.

Wintrobe, M.M. *Hematology: The Blossoming of a Science*. Lea & Febiger, Philadelphia, 1985.

Wailoo, K. *Drawing Blood: Technology and Disease Identity in Twentieth-Century America*. Johns Hopkins University Press, Baltimore, 1997.

Highways

The planning and use of highways, or motorways in the U.K., have profoundly altered the speed, scope, and meaning of individual transport as well as the face of the environment in many developed countries, especially during the second half of the twentieth century. Cars, which were originally designed as intraurban alternatives to horse-drawn carriages, could only be transformed into major carriers of regional and long-distance transportation with thoroughfares connecting urban centers. A highway is generally defined as a multilane road with separate lanes for each direction, separated by a median or crash barrier, to which access is limited both technically through entry and exit ramps and legally by only allowing cars and trucks to use these roads. Many countries require minimum speeds on their highways; all countries, with the exception of Germany and some U.S. states, limit the maximum speed. Often equipped with their own police, service stations, rest areas, and road crews, these roads have become institutions of their own, especially in those countries where they are run by separate public or private bodies. Research into surfacing, soil treatment, and the planning of interchanges, has contributed to the growth of these networks.

In both Europe and the U.S., calls for the creation of "car-only" roads grew louder after World War I. The overwhelming majority of roads were unpaved, and early motorists disliked sharing them with farmers, carriages, and livestock. The exclusion of the nonmotorized was at the heart of proposals to build car-only roads as promoted by middle class and upper middle class lobbies and local chambers of commerce. But the number of cars was still low and the status of the automobile predominantly that of a luxury. The Milan engineer Piero Puricelli introduced the idea of an "Autostrada dei Laghi" in 1922 to connect the northern Italian industrial centers. Only with the support of Mussolini was this fantastic idea realized. The Italian Fascists saw the autostrada project as a chance to portray themselves as energetic and in the vanguard of technological modernity. The first of Puricelli's roads was opened in 1924, and by 1935, 485 kilometers of autostrada had been built. They lacked separate carriageways, and drivers had to pay tolls on these roads.

Inspired by the Italian example, a coalition of industrialists and upper-class car drivers pushed unsuccessfully for similar roads called "autobahnen" in Germany. The lobby called itself "Hafraba" since the road was to connect the Hanseatic cities, Frankfurt, and Basle, and the lobby created technical blueprints. Toll roads were illegal in Weimar Germany, so the group lobbied unsuccessfully for the removal of this provision. The Nazi and Communist parties also opposed this restriction. Yet, after his assumption of power in 1933, Hitler grasped the opportunity to present his new government as proactive in the face of high unemployment, and he pushed the autobahn project against the advice of the national railway, the department of transportation, and the army,

which was opposed to building potential enemy targets. He created a separate government department for the Reichsautobahn, headed by the civil engineer Fritz Todt. The Nazi plans envisioned a network of 6,000 kilometers of roads—far more than demand—to be built by 600,000 laborers. When road building came to a halt in 1941, some 3,700 kilometers had been built by roughly 125,000 underpaid workers under harsh conditions. After the onset of World War II, Jews from Germany and the occupied countries were forced to work on these roads, shielded from public view. The exuberant propaganda of the autobahn, however, fashioned it into a symbiosis of technology and nature, thus belying the internal conflicts over road design between landscape architects and civil engineers. Through movies, radio, theater, paintings, and board games, Germans were told that "Adolf Hitler's roads" had been envisioned by the dictator during the 1920s. Another myth is the military and strategic role of the network during the war—motorways were of limited use to Nazi aggression.

When the Allied troops liberated Germany in 1945, civil engineers in the U.S. Army were astonished by the size of the autobahn network and were reminded of the lack of a similar system in the U.S. For the first four decades of the twentieth century, political conflict over taxes and financing had generally blocked interstate road construction in the highly federalized U.S. Local roads of varying standards did not extend into a transcontinental network. While "pleasure" roads such as the Merritt Parkway, the Mount Vernon Memorial Parkway, and New York City's Parkway System banned trucks, the Pennsylvania Turnpike was opened in 1940 as a labor-generating project using a railroad right-of-way. The federal Bureau of Public Roads, run by the technocrat Thomas MacDonald and aided by money from Congress, built or resurfaced over 145,000 kilometers of federally aided highways during the 1920s. The Bureau also sponsored and conducted road-building research, thus defining and homogenizing the standards for roads while ensuring the professional hegemony of the engineers. But not until the passage of the Interstate Highway Act in 1956 did these standards—including limited access and grade separation—become laws. After failing to pass the act the first time, the Eisenhower administration reintroduced the bill in Congress and advertised its ostensible military relevance. With remarkable speed, the U.S. set out to build the longest engineered structure ever, the Interstate Highway System. By 2000, some 89,400 kilometers of these roads spanned the U.S. European countries extended or commenced their highway programs in the 1950s, yet not with the same vigor as the U.S.

During the first decades of the Cold War, optimism about these roads abounded in the U.S. They were extolled as embodiments of a specifically American freedom to move and roam; extending the network was often seen as directly contributing to economic growth. The federal and state engineers controlling the construction program had few incentives to consider issues of urban renewal and social regeneration in their planning. As a result, urban interstates were often built through minority and lower-class neighborhoods, causing social displacement and environmental problems. The "freeway revolt" of the 1960s and 1970s rallied against the destruction of inner cities and ended the sole authority of engineers over the design process. A number of projects came to a halt, and roads increasingly came to be identified with pollution, the despoliation of nature, and suburban sprawl. This reevaluation of highways is true for Europe as well. When the Newbury Bypass in the U.K. was opened to traffic in 1998, it occurred in secrecy at 1:25 A.M. in order to forestall protests.

See also **Environmental Monitoring**

Thomas Zeller

Further Reading

Bortolotti, L. and De Luca, G. *Fascismo e Autostrada. Un caso di sintesi: La Firenze-Mare [Fascism and Motorways. A Case of Synthesis: The Motorway from Florence to the Sea]*. Franco Angeli, Milan, 1994.

Davis, T.F. Mount Vernon Memorial Highway: Changing Conceptions of an American Commemorative Landscape, in *Places of Commemoration: Search for Identity and Landscape Design*, Wolschke-Bulmahn, J., Ed. Dumbarton Oaks, Washington D.C., 2001, pp. 123–177.

Lewis, T. *Divided Highways: Building the Interstate Highways, Transforming American Life*. Penguin, New York, 1997.

Netzwerk Autobahn [Network Autobahn]. *WerkstattGeschichte* [thematic issue], 7, 21, 1998.

McShane, C. *Down the Asphalt Path: The Automobile and the American City*. Columbia University Press, New York, 1994.

Rose, M. *Interstate: Express Highway Politics*, revised edn. University of Tennessee Press, Knoxville, 1990.

Schütz, E. and Gruber, E. *Mythos Reichsautobahn: Bau und Inszenierung der "Straβen des Führers" 1933–1941 [The Myth of the Reichsautobahn: Building and Representing the Reichsautobahn, 1933–1941]*. Ch. Links, Berlin, 1996.

Seely, B. *Building the American Highway System: Engineers as Policy Makers*. Temple University Press, Philadelphia, 1987.

West, G. *The Technical Development of Roads in Britain*. Ashgate, Aldershot, 2000.

Zeller, T. The landscape's crown: landscape, perceptions, and modernizing effects of the German autobahn system, 1934 to 1941, in *Technologies of Landscape*, Nye, D., Ed. University of Massachusetts Press, Amherst, 1999, pp. 218–238.

Hip Replacement, *see* **Implants, Joints, and Stents**

Histology

Histology, simply defined, is the study of tissues. The term comes from the Greek words "histos," meaning web (or tissue), and "logos," meaning study; and the word histology first appeared in 1819. Without the microscope, there would be no modern field of histology; however Marie Francois Bichat, an anatomist and surgeon in Montpellier, France, defined 21 types of tissues without this technology. Other technologies used in histology pertain to preparation, preservation, and visualization of tissue samples.

The basic techniques of sectioning and staining used in the early part of the twentieth century depended on the ability of a tissue or cell to retain enough of its morphological integrity to be useful after being processed. One problem with preparing specimens was obtaining cuts of tissue thin enough for visualization without destruction. The microtome, a device for cutting specimens, developed synchronously in France and Germany. A handheld model by Nachet in 1890 was followed by a rotary microtome by Bausch and Lomb in 1901, and Leitz manufactured a sledge chain-driven machine in 1905. The basic principle involves the operation of a hand wheel that activates the advancement of a block of tissue embedded in wax toward a fixed knife blade that slices it into very thin pieces. Microtome knives were hand sharpened by histotechnologists until the 1960s when machines with disposable blades were introduced.

One challenge to early scientists was the death of cells after exposure to air and light: tissue scrapings placed on slides soon lost their shape and size. When tissues are removed from the body, they lose circulating nutrients and will deform unless treated with appropriate chemicals. Tissue had to be fixed, dehydrated, cleared, and infiltrated. During the first half of the century, each tissue was taken through a series of baths of formalin, alcohols, dioxane, and then paraffin. Formalin denatures the proteins so that they do not get damaged during the subsequent chemical baths. After rinsing, they were placed in 70 percent alcohol and successive baths of more concentrated alcohol until the application of xylene, a solvent for paraffin. Dioxane (diethyl dioxide alcohol) was used as the final dehydrant until the 1970s when its toxicity and pathogenicity were recognized.

After the tissue was cleared (alcohol removed) and made transparent, a material to support it was necessary. Liquid paraffin was embedded to infiltrate, support and enclose the specimens. In 1949, another embedding media, butyl metacrylate was introduced for the ultrathin sections used in electron microscopy. Celloidin, a nitrocellulose compound, was used in Europe instead of paraffin because it was considered superior with regard to support of tissues that were hard to infiltrate, such as bone or eyes. Carbowax, a water-soluble wax, was first used in the early 1960s. It took less time, but since it was hygroscopic, it required more care with regard to environmental moisture.

Until the 1970s, the technologist had to fabricate paper "boats" into which paraffin was poured to embed the prepared tissue. Some laboratories used plastic trays but until the introduction of an automated technology known as an embedding center (Tissue-Tek), the work was tedious and time consuming. After the paraffin cooled and solidified, the paraffin-containing specimen was cut on the microtome and mounted on a slide.

An alternative to the paraffin method of preparation is the cryostatic method, introduced in 1932 by Schultz-Brauns. This technology consisted of a refrigerated knife and microtome that could prepare the tissue at a temperature between –10 and –20°C. Its advantage was that so-called frozen specimens could be examined while the patient was still in surgery. If the tissue was pathological (usually cancerous), it could then be removed in the same surgical procedure. By the 1990s, the technique used was historadiography. Quickly-frozen tissues were dried, then prepared and photographed. The relative mass could be determined because there is a relationship between the film contrast and the various parts of the specimen.

Since cells are made from proteins, each subcellular organelle reacts to a dye by either staining to a color or not. The two most common stains are eosin, which stains the cytoplasm pink, and hematoxylin, a blue color used for nuclear material. The process of staining formerly required labor-intensive work in which a technologist applied stain to one slide at a time. The process took 12 hours from start to availability for examination in order to make a diagnosis. With the introduction of an automated system (Dako Autostainer) in the 1990s, 48 slides could be

processed in a period of two hours and with more than one stain. This equipment applies the stain, advances the slides, and dries them in a uniform process.

Preservation of live material is only temporary, but in certain cases one can observe both structure and function. The phase contrast microscope allowed for optimal visualization of difficult materials, but a challenge to progress in histology has been the inverse relationship between magnification and light. The power of a microscope depends on the wavelength of the light and the light-gathering capacity (numerical aperture) of the objective. Most histological work is performed with lenses of numerical apertures of 1.0 or less. If the wavelength of light is reduced, it is possible to increase the resolving power of a microscope to 0.1 micrometers with ultraviolet light. Smaller particles can be seen in dark field, but shape and dimension are not accurate. The challenge to increase both the light and the resolving power of the microscope resulted in a phase contrast microscope developed by Zeiss in the 1930s that was able to film a cell division. Such microscopes were not widespread until the 1950s. In 1955 improvements were made in the prism design, known as differential interference contrast (DIC), and this allowed visualization of living cells without staining.

In the 1960s a technology for diagnosing cervical cancer was developed by George Papanicolaou. Known as the PAP smear, the technique consisted of lightly scraping the cervical mucosa, spreading the sample on the slide, fixing it, all in the doctor's office, and then sending it to a laboratory for analysis. In 1996, a ThinPrep (Cytyc) test was approved by the U.S. Food and Drug Administration that collected the cells, rinsed them into a vial of preservative solution, and after filtration, applied them to a microscope slice uniformly. This technique was extended to non-gynecologic specimens such as fine needle aspirations and endoscopic brushings.

The halogen lamp, powered by a circuit board that prolonged its life, solved the light versus resolution relationship. Microscopes with multiple eyepieces allowed more than one person to view a specimen. By adding a tube to the viewing chamber and attaching another set of eyepieces, it was possible for two pathologists to view the same slide simultaneously (Olympus split microscope). Variations on this technology allowed for a class arranged in a circle to observe the same material. By the 1990s, technology had arrived at infinity-corrected optics; optical microscopy allowed investigation at the micron and submicron level.

Perhaps the most revolutionary development in twentieth century medicine has been the emergence of biotechnology. Monoclonal antibodies, immunohistochemistry, tumor markers, and flow cytometry depend on the ability to work with DNA and fluorescent labeling of organelles such as microtubules and endoplasmic reticulum. New diseases such as AIDS required new technologies to detect antibodies. By the end of the twentieth century, automation technology had affected and improved every division of histology departments in hospitals, from cytology and microbiology to pathology. One microscope could be accessorized with DIC, fluorescence, polarized light, phase contrast, and photomicrography using several film formats and digital image capture (Olympus Provis AX-70). The entire paradigm of a lengthy fixation time and preservation was modified because immunohistochemistry, electron microscopy, and molecular biology techniques required change. An autostainer can produce 2000 slides per hour, 100 at a time. NASA technology led to the development of an automated cellular information system (ACIS) using automated microscopy and computerized image analysis. The slide software captures hundreds of fields and projects them on a screen, and the observer selects those of interest and adjusts the magnification to a higher order. A quantitative score is then computed with regard to staining intensity and other parameters. The data is converted to a format suitable for export to a spreadsheet or database program. At the close of the century, fiber optics, digital cameras, and camcorders as well as computer software used in microscopy were used in programs for both students and continuing professional education.

LANA THOMPSON

Further Reading

Fredenburgh, J. and Myers, R.B. *The Effects of Fixation and Antigen Retrieval on Immunohistochemical Staining of Paraffin Sections*. Monograph. Richard-Allan Scientific, Kalamazoo, MI, 2000.

Long, E.R. Recent trends in pathology 1929–1963, in *A History of Pathology*. Dover, New York, 1965. Reprint of Williams & Wilkins 1928 edition.

Maximow, A. and Bloom, W. *A Textbook of Histology*. W.B. Saunders, Philadelphia, 1953.

Preece, A. *A Manual for Histologic Technicians*, 2nd edn. Little, Brown & Company, Boston, 1965.

Hormone Therapy

Hormones are substances formed in one organ of the body and transported in the blood to another

where they regulate and control physiological activity. The term hormone was coined in 1905 by William B. Hardy from the Greek meaning "arouse to activity." English physiologist Ernest H. Starling and Sir William Bayliss discovered the first hormone in 1902 and named it secretin. Chemically, hormones are divided into three classes: amino acid derivatives, peptides, and steroids. The purpose of hormone therapy is to provide or maintain physiological homeostasis with natural or synthetic hormone replacement.

Menopause

Perhaps the most familiar type of steroid hormone therapy is estrogen replacement therapy (ERT) for women whose ovaries have been removed surgically or women who have symptoms related to menopause. The first estrogen drug, under the trade name Premarin, was available in injectable form in 1942. The pill form gained popularity as the "medicalization" of menopause increased. The hormone was derived from the urine of pregnant mares. When the relationship between osteoporosis, hip fracture, and low estrogen levels was recognized, ERT was recommended for all women. The "baby-boomer" generation (those born between 1946 and 1964) was particularly receptive to ERT promotion, and by the end of the century one third of postmenopausal women in the U.S. were taking some kind of hormone replacement therapy.

In 1975, research indicated that unopposed estrogen raised the rate of endometrial cancers. Progesterone was added to estrogen, and the treatment became known as HRT (hormone replacement therapy). By the year 2000, esterified estrogens, synthetic estrogen from plant sources, and combinations with testosterone and progestins were prescribed as pills, transdermal skin patches, or intravaginal applicators. Studies early in the twenty-first century, however, determined that although these drugs were safer, they did not provide sufficient protection against osteoporosis or cardiovascular disease in women to warrant the risks of therapy.

Transgender Surgery

In 1930, Danish artist Einer Wegener underwent the first documented gender reassignment surgery and became Lily Elbe. Gender and sex, taboo topics for much of the century, became important research concerns in the 1960s. John Money began his longitudinal studies on children born with ambiguous genitalia (intersex individuals). Understanding how male and female sex hormones worked on the brain paved the way for more sophisticated hormone therapies before and after gender reassignment surgery.

Androgens are administered to biological females, and estrogens, progesterone, and testosterone-blocking agents (anti-androgens) are given to biological males, either orally, transdermally, or by injection, to stimulate anatomical changes. Orchiectomy (removal of the testicles) may mean that feminizing hormone dosages can be reduced. Unfortunately, follow-up studies revealed high morbidity and mortality rates, and the reasons for this are unknown. Although one can correlate hormone therapy with these rates, it is difficult to isolate causal factors because of the complex relationship between physiological and psychological processes.

Prostate Cancer

Hormone therapy may be initiated to treat prostate cancer following, or in addition to, surgery or radiation therapy. The hormones used to treat prostate cancers are also anti-androgens, substances antagonistic to the male hormones. Estrogen and diethylstilbestrol (DES) were used until the late 1970s when some synthetic compounds were introduced. In 1987 Zoladex, a drug that interferes with the production of testosterone, was tested to replace DES. Cyproterone acetate (a type of anti-androgen) and Estracyt or Emcyt (a combination of estradiol and nitrogen mustard) were found to be effective when estrogen therapy failed. A synthetic hormone to treat prostate cancer, Lupron, suppresses the formation of steroids in the body. Paradoxically, it has also been prescribed for the treatment of endometriosis in women.

Human Growth Disorder

Human growth hormone (HGH, somatotropin) was discovered by Herbert Evans in 1921, but it was not used in the U.S. until 1958 to treat certain kinds of dwarfism in children related to hormone deficiency. Stature deficits were virtually eliminated with this hormone, produced from human pituitary glands from cadavers. Unfortunately, the deaths of four children were traced to cadavers in which Creutzfeldt–Jacob disease infectious tissue was present (a new variant of Creutzfeldt–Jacob disease became known in the 1990s as the human form of mad cow disease), and in April 1985 all hormone production using cadavers ceased. In October of that year, a biosynthetic hormone, somatrem, was

produced using recombinant DNA technology. For thousands of children dependent on continued therapy for optimum growth, an unlimited supply of HGH could be produced. As with HRT for women, questions have been raised about the use and potential misuse of this hormone for human engineering.

Thyroid Deficiency

At the turn of the century, George Murray prepared an extract of thyroid from a sheep and injected it into a woman with myxedema, a severe form of thyroid deficiency, and the patient improved. Twenty-eight years later, he succeeded in making an oral preparation. Two hormones, thyroxine (T4), isolated in 1914, and triiodothyronine (T3), affect growth and metabolism, and both are controlled by thyrotropin, or thyroid stimulating hormone (TSH), produced by the pituitary gland. In 1910 the relationship between the thyroid gland and endemic goiter and iodine was discovered. When there is insufficient thyroid hormone, one cause of which is insufficient iodine in the diet, the body overproduces thyrotropin, causing enlargement of the thyroid gland; the condition is known as goiter. Replacement therapy suppresses thyrotropin, and the goiter shrinks. Despite the fact that thyroid and iodine supplements have been available for 50 years, endemic goiter and cretinism still occur worldwide.

Adrenal Hormones

Prior to the twentieth century, Addison's disease and its relationship with the adrenal glands were well known. But it was not until the 1930s that a great deal of interest was generated with regard to adrenocorticotropic hormone (ACTH). Deficiencies affected electrolyte balance, carbohydrate metabolism, hypoglycaemia, and sodium loss. It was postulated that two hormones were secreted by the adrenals: mineralocorticoids and glucocorticoids (salt and sugar hormones). Harvey Cushing described a syndrome in which too much ACTH was secreted. When the first extract of the adrenal gland was prepared in 1930, it was found to contain 28 different steroids, and five were biologically active. Adrenal cortex substance was of interest during World War II because of its effect in reducing the stress of oxygen deprivation. However, cortisone, the "anti-stress hormone," soon became the miracle drug of the twentieth century for treatment of arthritis and skin inflammations. In 1952, hydrocortisone, with fewer side effects, was developed using biosynthesis. By the 1960s, cortical steroids were being used for over 150 medical applications. Since the 1960s, technology has been directed at developing non-steroidal anti-inflammatory therapies.

Prostaglandins

In the mid-1930s, von Euler found that when the active substance of seminal fluid was injected into laboratory animals, it lowered their blood pressure. He named it prostaglandin. For the next thirty years his Swedish colleagues continued this work and discovered that prostaglandin was really four substances composed of fatty acids. This research led to the discovery of prostacyclin and thromboxane, hormones that have reciprocal actions on platelets. Prostacyclin has been used to treat circulatory problems. Flolan, a British product, is used to prevent blood from clotting in bypass machines.

Insulin

In 1921, Frederick Banting and Charles Best, two Canadian physiologists, were attempting to find a cure for diabetes mellitus. Their research on dogs involved removing the pancreas, waiting for symptoms of diabetes to develop, then injecting the animals with insulin. Previous experiments by others failed because the hormone degraded when taken orally. From the 1920s until the early 1980s, insulin was produced from the pancreases of pigs. In 1973, scientists at the Massachusetts Institute of Technology paved the way for synthetic insulin when they invented a technique for cloning DNA. In 1978 Genentech cloned the gene for synthetic insulin, and in 1982 the USDA approved an insulin derived from recombinant DNA called Humulin. Since that time, a disposable insulin pen has been developed to take the place of a syringe and vial. Other technologies include jet injectors, an external insulin pump, implantable insulin pumps, and insulin inhalers. Oral antidiabetic agents called sulfonylureas were developed in 1994. Another oral product, troglitazone, used concomitantly, was removed from the market in 1997 because of liver toxicity. In 2000, Eli Lilly, the original pharmaceutical company for the manufacture of insulin, announced a research partnership with Generex Biotechnology to develop another oral (buccal) form of insulin.

Negative Consequences

As previously noted, some of the technologies that have allayed suffering and prolonged life in the

twentieth century have also been misused. In the 1990s, synthetic HGH was marketed to athletes to increase body strength and to adults as the "anti-aging hormone." Many of these non-medical applications had disastrous consequences. Controversies with HRT continued as some women believed the "hot flashes" and sleep disorders they experienced beginning in perimenopause could only be relieved with hormone therapy, while others simply considered estrogen useful to maintain a youthful appearance. On the other hand, for the millions of children and adults who have benefited from insulin, there is no controversy about hormone therapy.

See also **Diabetes Mellitus; Contraception, Hormonal Methods and Surgery; Fertility, Human; Genetic Engineering, Applications**

LANA THOMPSON

Further Reading

Lock, S., Last, J.M. and Dunea, G., Eds. *The Oxford Companion to Medicine*. Oxford University Press, Oxford, 2001.

Maisel, A.Q. *The Hormone Quest*. Random House, New York, 1965.

Money, J. *Discussion in Endocrinology and Human Behaviour*. Oxford University, London, 1968.

Money, J. *Sex Errors of the Body: Dilemmas, Education, Counseling*. Johns Hopkins University Press, Baltimore, 1968.

Money, J. and Ehrhardt, A. *Man and Woman, Boy and Girl: The Differentiation and Dimorphism from Conception to Maturity*. Johns Hopkins University Press, Baltimore, 1972.

Pediatrics, in *Medical and Health Annual*. Encyclopedia Brittanica, Chicago, 1987.

Seaman, B. and Seaman, G. *Women and the Crisis in Sex Hormones*. Rawson Associates, New York, 1977.

Shames, K. and Shames, R. *Thyroid Power*. Harper Information, New York, 2001.

Hovercraft, Hydrofoils, and Hydroplanes

These three vehicle types combine aspects of flying and floating in hybrid marine craft capable of far higher speeds than traditional boat hulls. They accomplish this feat, valuable in commercial ferries, military vehicles, and racing boats, in very different ways.

Hovercraft

Hovercraft fly a few feet over water or land on a cushion of high-pressure air forced between the vehicle's hull and the surface by motorized (often turbine) fans. Though the air-cushion idea dates back more than 200 years, practical achievement of a viable carrier of people and goods came only in the mid-twentieth century, thanks to the inventive genius of radio engineer Christopher Cockerell in Britain and Jean Bertin in France. Vital to their success was C. H. Latimer-Needham's development in the late 1950s of an inflated "skirt" of rubber-like material hanging down several feet from the rim of the vessel to hold a deeper air cushion in place. Air cushion vehicles (ACV) or hovercraft saw their greatest promise in the 1960s as a variety of larger and faster designs were introduced, primarily in Britain.

The first full-sized hovercraft, the Saunders Roe-designed SR-N1, traveled across the English Channel in 1959. Experimental hovercraft ferry services began in 1962, and regular cross-Channel passenger and car ferry services began in 1964. Improved models soon reached speeds of 50 knots and could handle 10-foot (3-meter) waves. The SR-N4 two-ship class was introduced into regular car and passenger-carrying service in 1968. Demand for more capacity led to both being lengthened by 17 meters to increase their carrying capability to nearly 400 passengers and more than 50 cars in 1978–1979. These vehicles served as the world's largest hovercraft for three decades. But the growing cost of their operation and maintenance finally ended cross-Channel ferry hovercraft operations with the retirement of both vessels on 1 October 2000. They were replaced by other vessels including a growing fleet of Australian-built catamaran *Seacat* ships, first introduced in 1990 and able to carry more passengers and vehicles, as well as the opening of the Channel tunnel in the mid-1990s.

Hovercraft also serve a variety of rescue and military roles. Marines in several nations have taken hovercraft potential furthest, with over 100 craft in use in America and 250 in the former Soviet Union. The Russian "Bora" class, which entered service in 2000, displace more than 1,000 tons and carry guided missiles and a crew of 70. Other military hovercraft carry just one person for reconnaissance.

Hydrofoils

A hydrofoil is a metal wing that "flies" in water rather than air. This almost always means a boat or ship with fins attached to the light hull that can travel as either a normal vessel or up on the fins (hydrofoils) at much greater speed (up to 113 kilometers per hour (km/h)) due to the lack of hull friction with the water. Because water is far denser than air, the hydrofoils can be much smaller than

wings on an airplane yet still lift the hull when a minimum forward speed is achieved. Lift increases with speed, and a hydrofoil needs but half the power of a traditional ship. Most hydrofoils are ladder-like structures, enabling the marine craft to continue at high speed even in choppy seas as at least some of the hydrofoil structure remains under water. Hydrofoils are either surface piercing (intended to operate only partially submerged, and generally more stable) or fully submerged, with many variations of each. They can maneuver and take rough sea conditions more easily than traditional hull-borne vessels.

Though there were a number of nineteenth century hydrofoil ideas and experiments in both Europe and America, the first practical hydrofoils were designed in the late 1890s and early 1900s by Enrico Forlanini, an Italian engineer who worked on both advanced boats and airships. The first large-scale example, the HD-4, was designed by F. W. "Casey" Baldwin and a team backed by Alexander Graham Bell, and it achieved 113 km/h in tests in 1918. Parts of it, as well as a full-scale replica, are displayed in a Nova Scotia museum. Later developmental models were pursued into the late 1920s, but only HD-12, a 9-meter runabout, and HD-13, an outboard motor hydrofoil boat, were actually built, both in 1928

Hans von Schertel first began experimenting with hydrofoil craft in Germany in 1927 and had developed eight designs by 1936. Only in 1939 did the military first become interested in the potential of a hydrofoil boat. Various hydrofoils followed into late 1944, including one intended for torpedo attacks, one for coast defense, and another as a specialized landing craft. Schertel moved to Switzerland to continue his work after the war. In 1953 on Lake Maggiore connecting Switzerland and Italy, a 10-ton, 28-passenger von Schertal hydrofoil took 48 minutes to cross the lake (regular ferries took three hours). Using a similar design, in 1956 Carlo Rodriquez built several hydrofoils to carry passengers between Sicily and Italy and over the next four years more than a million people traveled by hydrofoil. In the 1950s and 1960s, the U.S. Army, Navy, and Marines all experimented with hydrofoil landing craft. Some of these were amphibians, and though successful, they were mechanically very complex and heavy for their limited payload capacities. The first U.S. Navy operational hydrofoil was the submarine chaser *High Point* of 1963, which served until 1978. A fleet of six (original plans called for a class of 30) Patrol, Hydrofoil, Missile (PHM) Pegasus-class navy vessels were built by Boeing Marine, based at Key West, Florida, and placed in service from 1975 to 1982. Utilized in drug interdiction among other missions, each displaced more than 250 tons and could reach 89 km/h on their hydrofoils. They were withdrawn in mid-1993 due to their high cost of operation.

Hydrofoils have operated as commercial ferries in all parts of the world, including several major urban harbors. Seven boats were part of the State Transit fleet of hydrofoils that operated between Sydney and Manly, Australia from 1965 to 1991 before being replaced by fast catamarans. Hong Kong hydrofoils have provided fast connections with Macao since well before both colonies were ceded to China.

Hydroplanes

Sometimes called planing craft or skimmers, hydroplane hulls resemble the simple surfboard in their basic function. By varying the shape of the bottom of a hull form (sometimes termed "stepped hulls"), a hydrofoil at speed can ride partially out of the water, thus decreasing drag and gaining greater speed. Such hull forms were, by the early twentieth century, being applied to maritime vehicles with somewhat similar principles later used in the hulls of flying boat aircraft.

The first serious discussion of such hull forms, however, dates back to the mid-nineteenth century. Abraham Morrison of Pennsylvania in 1837 patented a planing boat (though he did not call it that) with a concave hull bottom. A paper delivered in London the same year by noted engineer John Scott Russell noted that at speed such a craft "emerges wholly from the fluid and skims its surface." In the early 1850s Joseph Apsey proposed a steamship using a planing hull although the inefficient power plants of the day could not accomplish the task. In 1872 Charles Ramus suggested a wedge-hulled planing craft to the British Admiralty, which led to unsuccessful trials of a model vehicle. The problem again was the inadequate power generated by steam engines of the time—indeed, even rocket power was suggested! Finally, a Swiss inventor, M. Raoul Pictet, tested a model vessel on Lake Geneva in the early 1880s with a very modern concave hull form.

Sir John Thornycraft was one of several experimenters to resurrect the planing craft idea early in the twentieth century. In 1908 he proposed what he termed a hydroplane, featuring a two-wedge concave hull bottom that, driven with sufficient power, might largely ride above the surface with the aid of an air cushion between hull and water. At about

the same time, Comte de Lambert hit on an ideal combination when he developed a series of five floats with a structure to keep them aligned and featuring a blunt bow—with an engine driving a large airplane propeller mounted above the stern. This pioneering craft was slowly improved and one version entered passenger service on Lake Geneva in 1913, as did others on the River Nile. Modern versions of the light-hulled vehicles driven by huge caged propellers are now are widely used in the Florida Everglades and other swampy areas.

Hydroplane development in the postwar years centered primarily on constant attempts to achieve record-breaking speed rather than viable commercial services. Land vehicle racer Sir Malcolm Campbell developed a series of high-powered *Bluebird* hydroplane vessels that by 1939 had achieved better than 225 km/h in a closed course on England's Coniston Water. His son Donald died there in 1967 attempting to reach more than twice that speed when his jet-powered boat, *Bluebird K7*, broke up.

In the meantime in the U.S., Gold Cup Races beginning in Detroit in 1904 and later contests such as those running on Lake Washington near Seattle since 1950 and the Harmsworth International Trophy race, became the chief arenas for constant hydroplane hull and engine improvement. The *Slo-mo-shun* series of racing hydroplanes, for example, began development in the late 1930s. By 1950 the aircraft engine-powered *Slo-mo-shun IV* reached more than 257 km/h while the *Slo-mo-shun V* broke the 298 km/h barrier a year or so later.

Most racing hydroplanes since the late 1970s have been "propriders," so named as their driving propeller is partially submerged. The modern "three-point" (only two portions of the hull and part of its propeller are in the water at high speed) racing proprider largely rides on an air cushion and can be described as a kind of surface-effect vehicle.

CHRISTOPHER H. STERLING

Further Reading

Croome, A. *Hover Craft*, 4th edn. Hodder & Stoughton, London, 1984.

Cross, I. and O'Flaherty, C. *Hovercraft and Hoverports*. Pitman, London, 1975.

Gunston, W.T. *Hydrofoils and Hovercraft: New Vehicles for Sea and Land*. Doubleday, Garden City, NY, 1970.

Hynds, P. *Worldwide High Speed Ferries*. Conway Maritime Press, London, 1992.

Hogg, G. *The Hovercraft Story*. Abelard Schuman, London, 1970.

Jane's Surface Skimmers, Hovercraft and Hydrofoils. In 1994 became *Jane's High-Speed Marine Craft and Air-Cushion Vehicles*. Janes, London, 1968–date.

King, H.F. *Aeromarine Origins: The Beginnings of Marine Aircraft, Winged Hulls, Air-Cushion and Air-Lubricated Craft, Planing Boats and Hydrofoils*. Putnam, London, 1966.

McLeavy, R. *Hovercraft and Hydrofoils*. Blandford Press, London, 1976.

Parkin, J.H. *Bell and Baldwin: Their Development of Aerodromes and Hydrodromes at Baddeck, Nova Scotia*. University of Toronto Press, Toronto, 1964.

Watts, A.J. *A Sourcebook of Hydrofoils and Hovercraft*. Ward Lock, London, 1975.

Useful Websites

Classic Fast Ferries web magazine: http://www.classicfastferries.com/cff/

Hovercraft History and Museum:

Hydroplane and Racing Boat Museum: http://www.thunderboats.org/

Hydroplane history: http://powerboat.about.com/gi/dynamic/offsite.htm?site=http%3A%2F%2Fwww.lesliefield.com%2F%00

Hydroplane Racing: http://www.btinternet.com/~malcolm.mitchell/hydro/history.html

International Hydrofoil Society: http://www.foils.org/index.html

U.S. Hovercraft Society: http://www.flash.net/~bla77/ushs/TOC.htm

Hydroelectric Power Generation

It is estimated that about 50 percent of the economically exploitable hydroelectric resources, not including tidal resources, of North America and Western Europe have already been developed. Worldwide, however, the proportion is less than 15 percent.

The size of hydroelectric power plants covers an extremely wide range, from small plants of a few megawatts to large schemes such as Kariba in Zimbabwe, which comprises eight 125 megawatt generating sets. More recently, power stations such as Itaipu on the Parana River between Brazil and Paraguay in South America were built with a capacity of 12,600 megawatts, comprising eighteen generating sets each having a rated discharge of approximately 700 cubic meters per second.

Hydroelectric power has traditionally been regarded as an attractive option for power generation since fuel costs are zero; operating and maintenance costs are low; and plants have a long life—an economic life of 30 to 50 years for mechanical and electrical plant and 60 to 100 years for civil works is not unusual.

Small-scale hydropower schemes (typically less than 10 megawatts per site) utilize rivers, canals, and streams. Large-scale schemes generally include dams and storage reservoirs, with the option of pumped storage schemes to generate power to

match demand. Pumped storage schemes are however a net energy consumer, and should not be considered as renewable projects. Small hydroelectric installations are numerous in countries such as Scotland, South America, and China, for example, and may be operated by power generation companies or privately. Although some plants have been in service since the turn of the century, a considerable number of developments took place after 1945 and up to the mid-1970s, with a few, small, run-of-river developments having taken place since then. It is likely that the investment criteria that were applied in the later years of the twentieth century were more onerous than those set previously, and this has meant that new developments became more difficult to justify. The capital cost of "green field" (i.e., undeveloped and particularly unpolluted land) hydroelectric developments are higher than most alternative power generation schemes. Environmental concerns, for example over the Chinese government's undertaking of building the Three Gorges Dam, the largest hydroelectric project ever undertaken, must also be considered. In terms of a straight financial comparison, small hydroelectric plants are difficult to justify where the "competition" is a generating plant on a developed nationwide grid system. The existence of the National Grid in the U.K., for example, has allowed the exploitation of significant economies of scale in conventional thermal and later the combined-cycle generating plant.

Manufacturers' Developments

Although the field of hydroelectric engineering does not lend itself readily to the application of "standardization" techniques, most plant manufacturers have managed to use this approach to cover plants in the low output range. Most can offer packaged, or factory-assembled units, or plants that are broken down into major components, which can virtually be self-contained power stations. The accurate determination of the net head available for establishing the plant rating is fundamental. This involves deducting frictional head losses for the water intake and penstock (the conduit which carries the water to the turbine) and consideration of maximum and minimum head and tailwater levels with a statistical assessment of their frequency of occurrence.

Small and micro hydroelectric plants will generally only be viable if the civil works are simple. Even for sites where the head has already been developed (an existing dam, for example) it is unlikely that the civil engineering costs will be less than the cost of the plant.

The number, type and arrangement of generating sets will be decided upon at the feasibility study stage. The procedures for selecting the type of turbine and generator and for selecting synchronous speeds follow firmly established engineering principles, the objective ultimately being to employ the most cost-effective and appropriate overall arrangement to fulfill technical requirements. In collaboration with the civil engineer, the requirements for turbine setting (level of runner centerline relative to tailwater level to ensure cavitation-free operation), regulation issues (a function of the inertias of both the waterway and of the generating set), and the power station dimensions will be established.

A wide range of turbine types is available, each having distinct merits in certain fields and designed to cater for specific site conditions. The tubular turbine is a simple configuration of a propeller or Kaplan axial-flow reaction-type turbine suitable for low- to medium-head applications. A speed-increasing gear can be incorporated that enables a lower-cost high-speed generator to be employed. The cross-flow turbine is a partial admission radial impulse-type low-speed turbine. A speed-increasing gear or belt-driven arrangement is normally selected to permit cheaper high-speed generator designs to be used. The form of construction is simple, resulting in a lower-cost machine than the conventional types. The maximum efficiency is modest. The efficiency characteristic is, however, such that the machine can make efficient and therefore economical use of wide-ranging river flows.

The Francis turbine runner receives water under pressure from the spiral casing in a radial direction, discharging in an axial direction via a draft tube to the tailrace. For very low heads, the Francis turbine may be used in an "open flume" arrangement whereby the spiral casing is effectively replaced by a concrete forebay structure.

Pelton and Turgo turbines are impulse-type machines in which the penstock (potential energy) pressure is converted (via nozzles) into kinetic energy; the water jet impinges on the wheel and falls to the tailrace. Impulse-type turbines are relatively low speed and, therefore, except for small outputs, are not suited to the medium- to low-head range. Because the runner must be situated above maximum tailwater level to prevent "drowning," it is unavoidable that a proportion of the gross head cannot be utilized.

Two types of generator are used in hydroelectric installations: synchronous and asynchronous (or

induction) type. Synchronous generators are capable of independent isolated operation, synchronous speed and voltage control being maintained by the action of a governor (speed controller) as the load varies. An asynchronous generator relies upon the system to which it is connected to provide its magnetization current, and it operates at a power factor of 0.8 or less. This can be improved by using capacitors for "power-factor correction." While the turbine output must be automatically adjusted according to the availability of water, there is no need for governing. Synchronizing is uncomplicated as the generator circuit breaker is closed at or near synchronous speed. There are speed and output limitations for asynchronous machines, but these would probably not apply for the small hydro output range. The cost advantage diminishes with increasing output. The method of construction of asynchronous machines means that their inertia is lower than for synchronous machines. If this is an important consideration, however, it can be overcome by the use of a flywheel.

See also **Dams; Energy and Power; Electricity Generation and the Environment**

IAN BURDON

Further Reading

Greer, C. *Water Management in the Yellow River Basin of China*. University of Texas Press, Austin, TX, 1979.

Hay, D. *Hydroelectric Development in the United States, 1880–1940*. Edison Electric Institute, Washington D.C., 1991.

Schnitter, N.J. *A History of Dams: The Useful Pyramids*. Balkema, Rotterdam, 1994.

Wolf, D.E., Wolf, D.D. and Lowitt, R. *Big Dams and Other Dreams: The Six Companies Story*. University of Oklahoma Press, Norman, OK, 1996.

Worster, D. *Rivers of Empire: Water, Aridity, and the Growth of the American West*. Oxford University Press, New York, 1985.

I

Iconoscope

The iconoscope, a vacuum electronic camera tube, was the first television camera tube to clearly image a picture by all-electronic means. It was invented by Vladimir K. Zworykin, of the Radio Corporation of America (RCA), developed by his research group from 1929, and first field tested by RCA in an experimental television system in 1933. A similar tube known as the emitron was independently engineered by James McGee and his research group at Electric and Musical Industries from about 1932. The iconoscope and emitron all-electronic systems marked a shift away from the electromechanical television scanners that dated from the 1880s. Though no longer used, the iconoscope is the precursor to all television camera tubes used today.

The iconoscope has a vacuum-tight glass envelope. Inside there is a photosensitive mosaic, onto which an image is focused by a lens, and a high velocity electron beam that scans the mosaic line by line. The electrode configuration is illustrated in Figure 1. The electron gun that produces the electron beam comprises an oxide-coated cathode (C) mounted within a controlling electrode (G), which has a small hole through which the electron beam passes. The first anode (A) accelerates the

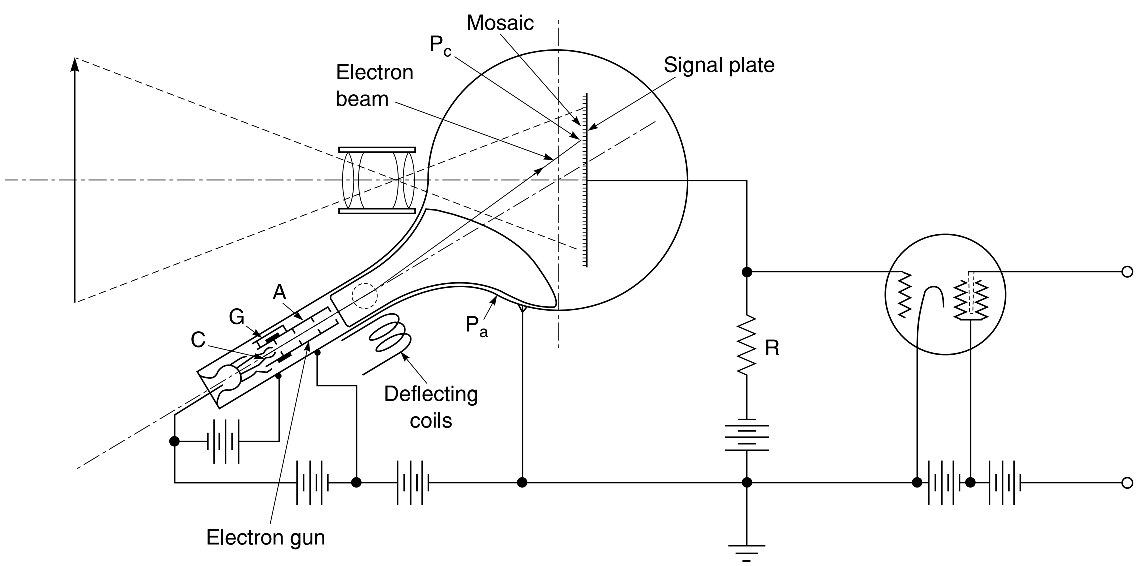


Figure 1. The iconoscope camera tube showing cathode C, control grid G, first anode A, second anode P_a, and the mosaic elements P_c.
[*Source: Journal of the IEEE, 73, 1933, p. 441.*]

electrons. The second anode (Pa), consisting of a metallic coating on the inside of the glass bulb, further accelerates the electrons to about one tenth the speed of light. The additional important function of this second electrode is electrostatically to focus the beam into a sharp spot on the photosensitive mosaic layer (Pc), which coats one side of a thin sheet of mica. A metallic coating, the signal plate, is deposited on the other side of the mica sheet. The mosaic consists of a very large number of minute, insulated, photosensitized silver globules, each of which release electrons under the effect of illumination (the photoelectric effect). Both frame and line-scanning deflecting coils enable the electron beam to scan a raster, or series of horizontal lines, on the mosaic's surface.

In operation, when the scanning electron beam strikes an unilluminated element of the mosaic, the high energy of the beam leads to the emission of secondary electrons that, if the emission is saturated, exceed the number of primary electrons of the beam. Since the element is insulated, it acquires an equilibrium positive charge, with respect to the anode, when the rate at which the secondary electrons leave the surface is equal to the rate at which the primary electrons strike the surface. If the mosaic is then illuminated (by an optical image of a scene or object), the mosaic elements emit photoelectrons, which are collected by the positively charged electrode. The mosaic elements consequently accumulate positive charges. These are returned to the equilibrium condition by the scanning primary electron beam, which restores the electrons lost by photoemission. Thus, during each scanning period (typically one twenty-fifth of a second), all the mosaic elements follow a cycle of positive charge acquisition by photoemission, and equilibrium charge restoration by primary electron beam scanning. Since each mosaic element is capacitively coupled to the signal plate, any variation of electric charge on the elements (caused by variation in light intensity) induces a corresponding variation of charge on the signal plate. The rate of change of this charge gives rise to the signal current.

A noteworthy feature of the iconoscope is that it incorporates the principle of charge storage, possibly the most important principle of early television engineering. With a single scanning cell, as used in the Nipkow disk–photocell arrangement (an invention of Paul Nipkow) and also Philo T. Farnsworth's image dissector tube, the cell must respond to light changes in $1/(Nn)$ of a second where N is the frame rate and n is the number of scanned elements. For N to equal ten frames per second and n to equal 10,000 elements (corresponding to a square image having a 100-line definition), the cell must react to a change of light flux in less than ten millionths of a second. But if the scanned mosaic of cells is used, each cell has 100,000 millionths of a second in which to react, provided that each cell is associated with a charge storage element. Theoretically, the maximum increase in sensitivity with charge storage is n, although in practice the very early iconoscopes only had an efficiency of about 5 percent of n.

The iconoscope is not free from defects. Apart from keystone distortion, which arises because the mosaic screen is scanned obliquely by the primary electron beam, the early workers found that the

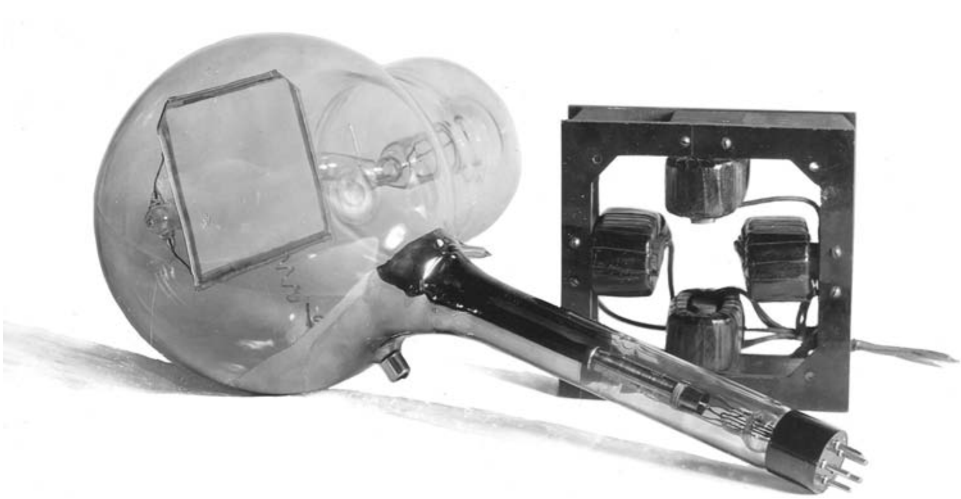


Figure 2. Iconoscope and deflecting electromagnets. [*Courtesy of the David Sarnoff Library*.]

picture signals tended to become submerged in great waves of spurious signals associated with the secondary electrons emitted. Correcting signals, which became known as "tilt" and "bend," had to be electronically generated and added to both the line and frame signals to annul the unwanted signals.

Electric and Musical Industries' emitron camera was used in the world's first all-electronic, public, high-definition television station, the London Station, which opened in 1936 at Alexandra Palace. The emitron cameras were light, portable, noiseless, and reliable. Their sensitivity—about 25 percent of the ideal—was sufficiently high to permit them to be employed for outside television broadcasts, including the 1937 Coronation, as well as for studio productions. Moreover, because the emitron cameras produced practically no lag in the picture, moving objects were reproduced clearly.

In 1934 Alan D Blumlein and James D McGee invented the ultimate solution—cathode potential stabilization (CPS)—to the problem of the spurious signals. In a CPS emitron, the primary beam approaches the mosaic in a decelerating field and strikes the surface with substantially zero energy. Hence no secondary emission of electrons can occur. Cathode potential stabilization has several advantages. First, the utilization of the primary photoemission is almost 100 percent efficient, thus increasing the sensitivity by an order of magnitude; second, shading (or spurious) signals are eliminated; third, the signal generated is closely proportional to the image brightness at all points (i.e., it has a photographic gamma of unity); and fourth, the signal level generated during the scan return time corresponds to "picture black" in the image.

These advantages became of much importance in the operation of signal generating tubes such as the CPS emitron, the image orthicon, the vidicon, the plumbicon, and every one of the Japanese photoconductive tubes up to the advent of the charge-coupled devices. The first public outside broadcasts with the then new CPS emitron cameras were from the Wembley Stadium and the Empire Pool during the fourteenth Olympiad held in London in 1948. The cameras were about 50 times more sensitive than the existing cameras and enabled what was proclaimed a wealth of detail and remarkable depth of field to be obtained.

See also **Television, Digital and High Definition Systems; Television, Electromechanical Systems; Television, Late Nineteenth and Early Twentieth Century Ideas**

RUSSELL W. BURNS

Further Reading

Abramson, A. *Zworykin, Pioneer of Television*. University of Illinois Press, Chicago, 1995.

Abramson, A. *The History of Television, 1880 to 1941*. McFarland, Jefferson, 1987.

Burns, R.W. *Television: An International History of the Formative Years*. Institution of Electrical Engineers, London, 1998.

Everson, G. *The Story of Television: The Life of Philo T. Farnsworth*. W.W. Norton, New York, 1949.

Fink, D.G. *Principles of Television Engineering*. McGraw-Hill, New York, 1940.

Maloff, I.G. and Epstein, D.W. *Electron Optics in Television*. McGraw-Hill, New York, 1938.

Zworykin, V.K., and Morton, G.A. *Television: The Electronics of Image Transmission*. Wiley, New York, 1940.

Immunological Technology

Immunization

At the start of the twentieth century, it was known that bacteria and some mysterious filterable agents called viruses could cause infectious diseases, and that people who recovered from a given infection often become specifically immune to that particular infection. There were only two vaccines known to be useful in preventing human disease, namely the widely used smallpox vaccine and the rarely used rabies vaccine of Louis Pasteur. It had been known for about 10 years that one molecular mechanism of specific immunity is the production of antibody, a protein produced usually in response to a foreign agent and able to bind specifically to that agent.

The smallpox vaccine evolved originally from the cowpox virus, which was used by Edward Jenner to immunize humans in 1796. During serial transmission from human to human, and later from rabbit to rabbit, it evolved into a distinctly different virus now called vaccinia. In the 1970s it played a central role in the first ever global eradication of a human disease: smallpox was officially declared eliminated from the human race in 1979.

Vaccinia was the prototype of many live viral vaccines, whose immunizing effects depend on their being able to cause an infection in the patient, albeit a very mild infection. Other live viral vaccines in common use today include those that immunize against measles, mumps, rubella, polio, varicella, and yellow fever.

These viral vaccines are all said to be attenuated, or of reduced virulence; that is, the infection that they cause is very mild compared with that caused by their wild viral ancestors. Attenuation of infective agents is usually achieved by very simple

methods, such as maintaining them for a long time in artificial culture or in an unusual animal host. Most of the attenuated viral vaccines produce strong, long-lasting protective immunity in most people after a single dose.

The only important attenuated bacterial vaccine is the BCG vaccine (Bacille Calmette–Guérin), which is a live, attenuated form of *Mycobacterius bovis*. It provides about 80 percent protection against tuberculosis in the U.K. but does not provide any demonstrable benefits when used in India.

Some vaccines are created merely by killing the infective agent by exposure to heat, alcohol, formaldehyde, or thiomersal. The usual pertussis (whooping cough) vaccines are of this sort, as is the killed polio vaccine. Others consist of purified single components of the infective agent. Examples are the vaccines against pneumococci, *Haemophilus influenzae* B, and meningitis C. The important vaccines against tetanus and diphtheria are made by purifying the very potent protein toxins that the bacterial agents produce and then exposing them to formaldehyde, which slightly alters the chemical structure of the toxin, making it nontoxic.

Some pathogens; for example, hepatitis B virus, are difficult to culture in a laboratory, and so the usual methods of creating a vaccine cannot be used. However, if only one protein is important for induction of the immunity, it is usually possible to transfer the gene that codes for it into a yeast or a bacterium that can easily be cultured. The required protein can then be recovered from the cultures and used for immunization. The current hepatitis B vaccine is of this type.

The net effect of vaccines in the twentieth century has been one of the most celebrated success stories of medicine. Beginning with the diphtheria vaccine in the early 1940s, many new vaccines have come into widespread use and the incidence of most childhood infections has fallen about 1000-fold. Smallpox has disappeared and polio is also well on its way towards extinction. On the other hand, there are still no affordable and effective vaccines against some of the world's biggest killers, such as malaria, schistosomiasis, trypanosomiasis, amoebic dysentery, and HIV.

Allergy

The same mechanisms that protect against infection sometimes swing into action in response to relatively harmless substances such as pollen, house dust mites, penicillin, aspirin, latex, wheat proteins, peanuts, fruits, shellfish, dyes, and nickel salts. These are known as allergens. Upon a second exposure to the same allergen, the body's immune system can overreact, causing disease rather than preventing it. This phenomenon is called allergy, a word that is usually restricted to adverse reactions known to be mediated by the immune system. When these results occur naturally, they can cause conditions such as hay fever, rhinitis, asthma, eczema, urticaria, celiac disease, and the life-threatening anaphylactic shock, in which blood pressure can fall dramatically and the bronchi can become severely constricted.

One of the great mysteries of the twentieth century is that allergies and asthma in particular were rare at the start of the century and progressively increased in prevalence in developed countries. By the end of the century about 20 percent of the population suffered from an allergy and about 14 percent of children in some areas suffered from asthma.

By about 1961 it was widely accepted that the prick test was one of the best methods of identifying which allergens were responsible for the most acute allergic reactions. A drop of the allergen solution is placed on the skin and a trace is introduced into the epidermis by pricking it gently with a lancet. If the patient is allergic, a swollen, pale, edematous weal appears within 15 minutes at the site of pricking, surrounded by a reddened flare region that itches intensely. This test helps to decide which allergens the patient ought to avoid.

Management of allergic patients has been greatly improved by the use of drugs such as sodium cromoglycate that prevent histamine from being released from mast cells, and by antihistamines that prevent the histamine from exerting its inflammatory effects. For asthma, a slow-acting but long-lasting benefit is conferred by inhaled steroids such as Beclomethasone, also known as Vanceril or Beclovent, while an acute attack can be treated with β-adrenergic agonists such as Salbutamol, which relaxes the smooth muscle of the bronchi. The main therapy for anaphylactic shock is injected adrenalin, also known as epinephrine.

Desensitization to insect venoms and inhaled allergens such as pollen can often be achieved by giving the patient a long course of immunotherapy with very small but increasing doses of the allergen, but this is seldom successful with food allergens.

Antibodies

Injection of antitetanus antibodies into patients at risk of tetanus was introduced in 1914 and had an

immediate, major impact on the number of deaths from tetanus among those wounded on the battlefield. Antibodies produced by a patient during an infection are sometimes used to identify the infective agent, especially in virus infections, and they have been used since 1900 to detect differences among different strains of bacteria and different blood cells. Many species of salmonella are distinguished largely by their reaction with antibodies, and blood typing for transfusion purposes depends largely on observing the reaction of the red cells with antibody.

Since antibodies can be made that react specifically with almost any soluble substance of our choice, they provide a convenient type of reagent that can be used to assay many different substances or to locate them in living tissues. Popular techniques based on this idea include radioimmunoassays and enzyme-linked immunosorbent assays (ELISA). The simplest, direct immunofluorescence techniques involve covalently coupling an antibody with a fluorochrome to produce a staining reagent that has the extraordinary specificity of antibodies as well as the extraordinary ease of detection of fluorochromes. It may be used to demonstrate the precise location of a particular substance within a microscopic section of a biological tissue.

All these techniques originally suffered from the fact that it was impossible to make exactly the same antibody twice, and so any assay had to be revalidated and recalibrated when a given batch of antibody was exhausted. But in 1975 the immunologists Georges Kohler and Cesar Milstein showed how to create a potentially immortal clone of cells that would continue to make the same antibody for as long as required, thereby allowing many different researchers to use an identical product. This monoclonal antibody also had some advantages arising from the fact that all of its molecules had very nearly the same structure.

Autoimmunity

In 1945 a case of autoimmune hemolytic anemia was described by Coombs, Mourant, and Race. They showed that the patient's anemia was caused by an antibody that reacted with his own red blood cells, triggering their destruction. Since then many human diseases have been found to be associated with immunity to one's own tissues. These diseases include rheumatoid arthritis, pernicious anemia, insulin-dependent diabetes, Hashimoto's thyroiditis, thyrotoxicosis, idiopathic Addison's disease, myasthenia gravis, psoriasis, vitiligo, systemic lupus erythematosus, and many others.

Steroids and immunosuppressive drugs have helped to control these conditions, and some therapeutic antibodies against inflammatory mediators have initially showed some promise, but a cure is not yet available.

Transplantation Immunology

In 1900 the ABO blood groups were discovered by Karl Landsteiner, and this made it possible to match blood donors with recipients and therefore transfuse blood safely. This may be regarded as the start of transplantation. Unfortunately, almost all human cells other than red blood cells bear on their surface not only the ABO substances but also a set of molecules known as histocompatibility molecules, which differ greatly between individuals and result in strong immune reactions against most transplanted tissues.

The problem can be partly alleviated by matching donor with recipient, first tried in 1966, but a complete match is almost impossible except between monozygotic twins.

Organ transplantation, however, became an important and useful treatment with the development of cytotoxic drugs that suppress the immune response. The main such drug is cyclosporin A, first used in 1978. The prevention of immune-mediated rejection, however, is not quite complete. A low-grade immune reaction still may limit the survival of the graft, and the drugs have several unfortunate side effects, including the suppression of desirable protective immune responses. Some see therapeutic cloning as a potential long-term solution to the problem of transplant rejection. Others hope to find an effective method of inducing specific tolerance to the donor tissue.

See also **Organ Transplantation**

IAN C. MCKAY

Further Reading

Elliman, D.A.C. and Bedford, H.E. MMR vaccine—worries are not justified. *Arch. Dis. Childhood*, 85, 71–274, 2001. A concise review of the evidence concerning safety of the MMR vaccine.

Voltaire (*Lettres*. 1733). Describes the history of variolation and the incidence and effects of smallpox in the early eighteenth century.

Useful Websites

Centers for Disease Control and Prevention. General recommendations on immunization: recommendations of the Advisory Committee on Immunization Practices and the American Academy of Family Physicians.

Morbidity and Mortality Weekly Report (MMWR), 51, RR-2), 2002. All *MMWR* references are available on the Internet at: http://www.cdc.gov/mmwr
Information on general recommendations on immunization, and information about the National Immunization Program: http://www.cdc.gov/
World Health Organisation, Information on Vaccines and Immunizations: http://www.who.int/m/topics/vaccines_immunizations/en/index.html
Public Health Laboratory Service, U.K.: http://www.phls.co.uk/
Web pages on immunization maintained by the author of the above article: http://www.immunol.com

Implants, Joints, and Stents

Joint replacement, particularly for the hip and knee, has become a common surgical procedure for the treatment of joints affected by arthritis. Joint replacement replaces the surfaces of the natural joint and leads to restoration of joint function and pain relief.

The hip consists of the articulation between the spherical femoral head and the cup-shaped acetabulum. The first total hip replacement was performed in 1938 by Philip Wiles at the Middlesex Hospital in London. The design, made from stainless steel, consisted of a spherical femoral component attached to a bolt passing down the femur and a cup-shaped acetabular part that was secured by screws. However, it was not until the 1950s and 1960s that the idea of joint replacement became possible through the pioneering work of Kenneth McKee (Norwich, U.K.) and John Charnley (Wrightington, U.K.). McKee had designed a hip replacement, similar to the one used by Wiles, with a stainless steel ball and socket secured by screws. However, the results of replacements were poor because of inadequate fixation.

Charnley developed the idea of using dissimilar materials to create a low friction implant. He used a stainless steel femoral component and plastic socket made from polytetrafluorethylene (PTFE), more commonly know as Teflon. Charnley also introduced the idea of using large amounts of acrylic cement to fix the implant. The cement was used as a grout, relying on the mechanical fit rather than a glue. Over 300 hip replacements were undertaken by Charnley before it was realized in 1962 that PTFE was not a suitable material for hip replacement. Studies found that high wear rates resulted in a severe tissue reaction. Later in 1962, Harry Craven, the technician at Charnley's biomechanical laboratory, obtained a new plastic known as high-molecular weight polyethylene (HMWPE). At first Charnley was dismissive of the new material, but Craven tested it on a wear machine and found that it had much lower wear rates than PTFE and appeared to be a better material for hip replacement. Charnley began implanting the new metal-on-plastic hips in November 1962, and this is the basis of the metal-on-plastic hip implant that was most commonly used up to the twenty-first century. The metal component, however, was often a cobalt chrome molybdenum alloy (CoCrMo), and the plastic was ultra-high molecular weight polyethylene (UHMWPE) (see Figure 3).

Metal-on-polymer hip replacements can be expected to last for at least 15 years, but the wear particles developed from the metal articulating against the polymer leads to wear debris, which can cause osteolysis. Osteolysis is a tissue reaction that causes bone resorption and loosening of the implant, and the problem has led to growing interest in alternative biomaterials for hip replacements. Metal-on-metal (CoCrMo) and ceramic-on-ceramic (alumina) are considered to lead to less wear debris and to avoid the problem of osteolysis.

The knee consists of the articulation between the femur and the tibia; the two bones being separated by fibrous cartilage called the meniscus. In addition

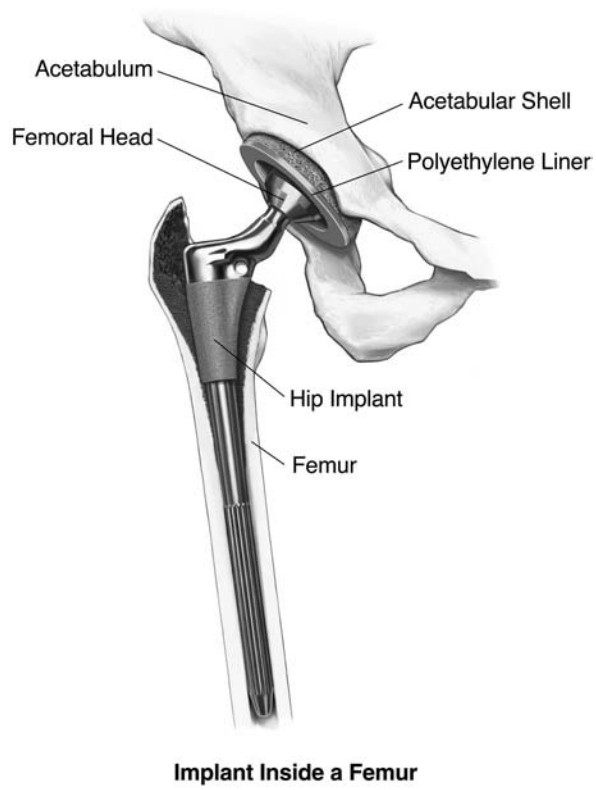


Figure 3. A total hip replacement showing the femoral and acetabular components.
[*Reproduced with kind permission from Zimmer Inc., U.S.*]

the patella (knee cap) articulates with the femur. The development of knee replacements followed a similar trend to that of hip replacements in which early attempts failed, mainly because of loosening. In the early 1950s, hinged knee replacements were developed by Wallidus (1951) and Shiers (1953). These two implants consisted of a stainless steel hinge, but they required large bone resection, and there were high failure rates associated with them. Modern-day knee replacement took off after the success of the Charnley metal-on-polymer hip replacement. In 1969, Gunston, who worked with Charnley at Wrightington, developed a metal-on-polymer knee replacement that was fixed with acrylic cement. The implant consisted of a stainless steel femoral component and a HMWPE tibial component.

By the end of the twentieth century, there were various designs of knee replacement available that consisted of femoral, tibial, and/or patellar components. The implant can be: (1) unconstrained; (2) semi-constrained; or (3) fully constrained, depending on whether the knee ligaments are present. Implants are also divided according to the proportion of the knee replaced: (1) unicompartmental to replace either the medial or lateral compartment; (2) bicompartmental to replace both medial and lateral components; (3) tricompartmental to replace medial and lateral components and the patella. Present-day knee replacement (see Figure 4) consists of a metal femoral component (CoCrMo) that has a stem cemented into the femur, a tibial stem of titanium alloy cemented into the tibia, a tibial tray (UHMWPE) that fits to the tibial stem, and a patellar component (also of UHMWPE).

The fixation, or setting, of fractures has been aided by the use of screws and pins. The use of screws in bone began around the late 1840s, and today bone screws are commonly used either alone to treat fractures or in conjunction with plates or intramedullary nails. They are generally made from stainless steel or titanium alloy. The screws can either be self-tapping (the screw cuts a thread as it is inserted) or non-self-tapping (requires a tapped hole). Two main types of screws are used to get good purchase in the two types of bone: cortical (small threads) and cancellous (large threads). The holding power of a bone screw depends on a number of factors including diameter, length of thread engagement, and thread geometry. Hence, there are many different designs, lengths, and diameters of screw available for fracture fixation. Screws can also be cannulated (have a hole down the center) for use with wires and pins.

Pins or wires are also used in fracture fixation. The use of wires dates to the 1770s. Straight wires are known as Steinmann pins, while wires of diameter less than 2.4 millimeters are known as Kirschner wires. Wires and pins are primarily used to hold bone fragments together, either temporarily or permanently. They can also be used to guide large screws or intramedullary nails during insertion. The pins have a sharpened tip designed to penetrate easily through the bone.

A stent is a device used to maintain an orifice or cavity in the body. Stents are mainly used in vascular diseases to maintain blood flow, although they can be used to maintain a passage in other sites such as the urinary duct or for a bronchial obstruction. The most common type of stent used for coronary artery disease is made from fine stainless steel wire; the device has a lattice appearance resembling chicken wire. In angioplasty procedures, the stent is inserted into the

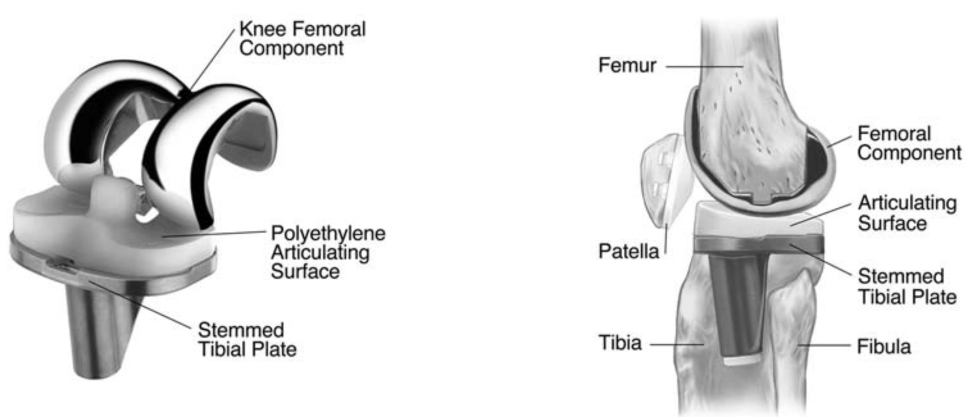


Figure 4. A total knee replacement showing the femoral, tibial, and patellar components. [*Reproduced with kind permission from Zimmer Inc., U.S.*]

artery on a balloon catheter. The balloon is inflated, causing the stent to expand and press against the inside of the artery. Once the balloon has been deflated and removed, the stent remains in place keeping the artery open. Other stents can be made from a variety of materials: nitinol (a nickel–titanium alloy), polymers, elastomers, and biodegradable materials.

See also **Cardiovascular Surgery and Implants**

DUNCAN E.T. SHEPHERD

Further Reading

Chandran, K.B., Burg K.J.L. and Shalaby S.W. Soft tissue replacements, in *The Biomedical Engineering Handbook*, Bronzino, J.D., Ed. CRC Press, Boca Raton, FL, 2000.

Dowson, D. New joints for the millennium: wear control in total replacement hip joints. *Proc. Inst. Mech. Eng: Part H, J. of Eng. Med.*, 215, 335–358, 2001.

Dowson, D. and Wright, V. *An Introduction to the Biomechanics of Joints and Joint Replacement*. Mechanical Engineering Publications, London, 1981.

Park S.-H., Llinás, G.V.K. and Keller J.C. Hard tissue replacements. in *The Biomedical Engineering Handbook*, Bronzino, J.D., Ed. CRC Press, Boca Raton, FL, 2000.

Waugh, W. *John Charnley: the Man and the Hip*. Springer Verlag, London, 1990.

Park, J. and Lakes, R. *Biomaterials: an Introduction*. Plenum, New York, 1992.

Industrial Gases

While the gases that are now commonly referred to as "industrial," namely oxygen, hydrogen, carbon dioxide, and nitrogen, were not fully understood until the nineteenth century, scientists in the twentieth century moved rapidly to utilize the knowledge. Driven largely by the demands of manufacturing industries in North America and Western Europe, rapid improvements in the technology of production and storage of industrial gases drove what has become a multibillion dollar business, valued at $34 billion in 2000. At the start of the twenty-first century, industrial gases underpin nearly every aspect of the global economy, from agriculture, welding, metal manufacturing and processing, refrigerants, enhanced oil recovery, food and beverage processing, electronic component manufacturing, to rocket propulsion. Oxygen for metal manufacturing is the largest volume market, with chemical processing and electronics using significant volumes of hydrogen and lower volumes of specialty gases such as argon.

Until at least the fifteenth century, gases were thought of inclusively as "air," part of the alchemical construction of the world. It was not until the seventeenth and eighteenth centuries that properties and characteristics unique to individual gases were recognized. Once one gas was finally isolated, a cavalcade of similar discoveries followed. In 1754, for example, English aristocrat and chemist Henry Cavendish "discovered" hydrogen, in 1756 English chemist Joseph Black discovered that carbon dioxide was a constituent of carbonate rocks, in 1772 Swede Carl Wilhelm Scheele discovered oxygen's properties, Englishman Joseph Priestley isolated oxygen (which he called dephlogisticated air) by 1774, and by 1784, Frenchman Antoine Laurent Lavoisier came up with a method of decomposing water into constituent elements, which he named hydrogen and oxygen, by passing water vapor over hot charcoal. Lavoisier was one of the first quantitative chemists, and showed that water was composed of two-thirds hydrogen and one-third oxygen.

These earliest discoveries highlighted the challenges ahead for harnessing the benefits of these gases: how to isolate and store them. Hydrogen, known for its "lighter than air" qualities as early as 1774, was first extracted from water using electrolysis by William Nicholson and Anthony Carlisle in 1800, but the high cost of electricity proved to make this an expensive method. As demand increased for hydrogen's use in airships, from zeppelins to military observation balloons in the 1930s, more economic means of extraction appeared. Since the 1920s, hydrogen has been produced by liquefaction of natural gas, partial oxidation of heavy oil, or gasification of coal.

Oxygen was first used in medicine (as an anesthetic or ventilator) and for "limelight," a theatrical lighting method from burning oxygen and hydrogen together. Early oxygen-using equipment could function with low-purity oxygen in small amounts, acquired by using several different chemical and heating processes that would break the oxygen into molecules, which was then compressed and sold in cylinders. By 1902, a system of rectifying "liquid air" pioneered by German Carl von Linde produced oxygen up to 99 percent pure.

In the early nineteenth century, nitrogen fertilizers were obtained from discoveries of immense bat guano deposits and caliche (nitrate-bearing rocks) in South America. As threats of famine loomed at the close of the nineteenth century, the need for agricultural fertilizers drove early production of nitrogen from the atmosphere. While "liquid air" contained nitrogen, it was difficult to separate from oxygen. Three solutions were developed in Europe: the cyanamide process (c.1900), which involved passing steam over certain carbides to form calcium cyanamide; the electric arc process (c.1903), which

imitated lightning discharges to isolate nitrogen from air; and the Haber process, created by German Fritz Haber in 1904 and later developed into an industrial process by Carl Bosch, in which nitrogen is reacted with hydrogen to form ammonia.

By the twentieth century, carbon dioxide was extracted from many natural sources, especially cracks in the earth's crust due to volcanic activity, and as a byproduct of limekiln operations and synthetic ammonia production. Carbonic acid was used in bottled soft drinks, cooling, and in dry ice. Combined with sodium bicarbonate and ammonium sulfate it also created a foam, which deployed from a pressurized canister, became the fire extinguisher (carbon dioxide will not support combustion, and foam application ensures the gas does not quickly disperse).

Acetylene, discovered in 1836 and used in home and street lighting, was not produced industrially from calcium carbide until 1892. In 1901 Charles Picard and Edmond Fouché separately invented the oxyacetylene lamp, now widely used in arc welding, by combining acetylene with oxygen to produce an intense heat. Argon was used as an inert gas for electric lights, neon was used in lighted signs by 1913, and helium was used in balloons, dirigibles, welding, and medicine, and also mixed with oxygen for compression into cylinders for divers (the first practical compressed air diving apparatus was produced in 1925).

After World War II, applications for industrial gases expanded, using combinations of the basic four with lesser-known gases, as with the oxyacetylene lamp. Because acetylene was highly volatile in its compressed (and saleable) form, several innovations in transportation of the gas became necessary, with acetylene transported in pressurized steel cylinders.

However it was the liquefaction of gases—bringing a gas to a liquid state by intense pressure followed by cooling expansion—that was critical to the transport and application of these gases. Building on compression technology invented in the late nineteenth century, the science of cryogenics was the major contribution to liquefying gases. The method of cryogenic separation or distillation developed by German Carl von Linde in 1895 for liquefied air, involved dropping the temperature of the air to below –140°C and increasing pressure to eight to ten times that of the atmosphere. As the liquid air is boiled, gases are boiled off at different boiling temperatures: vaporized nitrogen first, then argon, then oxygen. By 1908, all the gases had been separated using several different cryogenic machines, leading to a boost in the availability of these liquid gases, and increasing demand as subsequent uses for the gases developed. In the post-World War II years, oxygen was produced in large quantities, thanks to the Linde–Frankl process for liquefying air (developed in 1928–1929), which ultimately gave way in the 1960s to the molecular sieve or pressure swing adsorption—a method that removes carbon dioxide from air. The abundance of oxygen also contributed to great advances in medicine—namely anesthesia and respiratory support, as well as facilitating high-altitude flying.

The growth of industrial use of oxygen and acetylene meant that pressure vessels were needed for transportation and storage. One of the earliest storage devices for liquefied gas was a double-walled vessel with a vacuum in the annular space, known as the "Dewar flask," invented by Sir James Dewar in 1892. Building on this early technology, ultrahigh vacuum pumps, as well as aluminum and foil matting used as insulation, contributed to advances in the storage and transport of highly pressured gases. In 1930, the first vacuum-insulated trains and trucks carried refrigerated, liquefied gases in the U.S., and by the 1950s, pipelines carried gas—with the largest in France—transporting oxygen, nitrogen and hydrogen from several different plants. The first generation of compressed gas cylinders (1902–1930) used carbon steel cylinders. Problems with rupturing led to the development of quenched and tempered alloy steel cylinders.

By the 1960s, oxygen was being used by steel manufacturers to enhance combustion in furnaces, and by the 1970s, nitrogen was widely employed as inert packaging for food preservation, as well as freeze-drying of food and heat treatment of metals. By the 1960s liquid hydrogen was used in rocket fuel and as a coolant for superconductors. By the 1980s, semiconductors were big customers for bulk industrial gases such as oxygen, argon, and hydrogen used as carrier gases for epitaxial growth, with specialist applications demanding higher quality gas (for example high-purity silane as a dopant, chlorine as an etchant). Cryosurgery using liquid nitrogen contributed to advances in medicine, including fertility treatments with frozen embryos, blood bank storage, and organ transplantation.

Industries from the automotive industry to steel making rely heavily on industrial gases. Hydrogen is used to cool alternators in power plants, and high-pressure cylinder gases are used in hospitals, small welding businesses, and in fire extinguishers. Liquefied carbon dioxide is used for dry cleaning textiles and as an industrial solvent, for example degreasing machine parts, coffee and tea decaffei-

nation, and extracting essential oils and medicinal compounds from plants. While cryogenic distillation is still the favored form of production for industrial gases, some industries have shifted to less expensive noncryogenic methods. Pressure swing adsorption, for example, pumps pressurized air into either a molecular sieve, or an adsorptive agent, which removes the "unwanted" gas from the air stream, leaving the desired gas behind (for example, removing carbon dioxide from air). While the end product is not as pure as cryogenic products, engineers are perfecting the method. Noncryogenic separation techniques such as pressure swing adsorption have made on-site production more affordable. On-site production, primarily of oxygen, reduces distribution costs for remote locations, and guarantees essential supply, for example for hospitals.

One of the byproducts of the expansive use of industrial gas is an increase in undesirable environmental pollutants—contributing to the "greenhouse effect" and an overabundance of nitrates from agriculture application. Subsequently, government controls worldwide have led the gas industry to revamp some of its distribution and application. Hydrogen is likely to be the gas of the future, employed in "green" fuel cell technology, and glass and steel manufacturers are reducing nitrous dioxide emissions by mixing oxygen with coal.

See also **Cryogenics: Liquefaction of Gases; Nitrogen Fixation**

LOLLY MERRELL

Further Reading

Almquivst, E. *History of Industrial Gases*. Kluwer Academic/Plenum, New York, 2003.

Anonymous. Something in the air. *Process Engineering*, 32, Jan. 2000.

Downie, N. *Industrial Gases*. Kluwer Academic, Dordrecht, 1996.

Gardner, J. One hundred years of commercial oxygen production. *BOC Technology Magazine*, 5, 3–15, 1986.

Kroesche, A. Out of thin air: due to technology and environmental concerns, industrial gases are seeing good growth opportunities ahead. *Chem. Mark. Rep.*, 243, 17, 1993.

Scurlock, R.G., Ed. *History and Origins of Cryogenics*. Oxford University Press, New York, 1992.

Wett, Ted. Technology galore. *Chem. Mark. Rep.*, 240, 24, 13–14, 1991.

Information Theory

Information theory, also known originally as the mathematical theory of communication, was first explicitly formulated during the mid-twentieth century. Almost immediately it became a foundation; first, for the more systematic design and utilization of numerous telecommunication and information technologies; and second, for resolving a paradox in thermodynamics. Finally, information theory has contributed to new interpretations of a wide range of biological and cultural phenomena, from organic physiology and genetics to cognitive behavior, human language, economics, and political decision making.

Reflecting the symbiosis between theory and practice typical of twentieth century technology, technical issues in early telegraphy and telephony gave rise to a proto-information theory developed by Harry Nyquist at Bell Labs in 1924 and Ralph Hartley, also at Bell Labs, in 1928. This theory in turn contributed to advances in telecommunications, which stimulated the development of information theory *per se* by Claude Shannon and Warren Weaver, in their book *The Mathematical Theory of Communication* published in 1949. As articulated by Claude Shannon, a Bell Labs researcher, the technical concept of information is defined by the probability of a specific message or signal being picked out from a number of possibilities and transmitted from A to B. Information in this sense is mathematically quantifiable. The amount of information, I, conveyed by signal, S, is inversely related to its probability, P. That is, the more improbable a message, the more information it contains. To facilitate the mathematical analysis of messages, the measure is conveniently defined as $I = \log^2 1/P(S)$, and is named a binary digit or "bit" for short. Thus in the simplest case of a two-state signal (1 or 0, corresponding to on or off in electronic circuits), with equal probability for each state, the transmission of either state as the code for a message would convey one bit of information. The theory of information opened up by this conceptual analysis has become the basis for constructing and analyzing digital computational devices and a whole range of information technologies (i.e., technologies including telecommunications and data processing), from telephones to computer networks.

As was noticed early on by Leo Szilard in his classic 1929 paper on Maxwell's Demon and then later by Denis Gabor in ideas on light and information (1951), information in this technical sense is the opposite of entropy. According to the Second Law of Thermodynamics, the entropy or disorder of any closed system tends naturally to increase, so that the probability of energy differentials within it approaches 0. A heat engine, for instance, depends on the existence of an improb-

able or anti-entropic energy differential, created typically by energy input from a confined heating source, so that energy then flows in accord with the Second Law from one part of the system to another (as in the Carnot cycle of a steam engine). But in a late nineteenth century thought experiment, the physicist James Clerk Maxwell posed a paradox. Imagine, Maxwell suggested, a closed container with a gas at equilibrium temperature or high entropy; that is, without any detectable energy differential. In such a case, the standard distribution curve will nevertheless dictate that some gas molecules possess slightly higher energy states than others. Introduce into the container a partition with a small door operated by a "demon" that lets randomly higher energy molecules move one way and randomly lower energy molecules move the other. The demon will thus tend gradually to create an energy differential in the system, thus reducing entropy and creating a potential heat engine, without any energy input. Before the formulation of information theory, Maxwell's demon seemed to be getting something for nothing, whereas in fact it is introducing information (about the energy of molecules) into the system.

Information thus functions like negative entropy or "negentropy" (a term coined by Erwin Schrödinger). Expanding on the connections between information theory and thermodynamics, Norbert Wiener developed the closely related theory of cybernetics (1948) to analyze information as a means for control and communication in both animals and machines. In Wiener's theory, Maxwell's demon becomes a *kubernetes* (the Greek word for steersman) who utilizes some small portion of a system output as information and energy feedback to regulate itself in order to attain or maintain a predefined state.

To no machine has information theory been applied with more intensity than the digital computer. Indeed, as one computer scientist, Goldstine, noted in 1972, despite the implications of its name, a computer "does not just operate on numbers; rather, it transforms information and communicates it." With a computer the simple transmission of encoded information, as in telecommunications networks, is transformed into a complex storing and processing of the code so as to sort, compound, interrelate, or unpack messages in ways useful to machine or human operations. One may think of advanced computers as composed of millions of differentially programmed Maxwell demons embedded in the silicon of integrated circuits so as to yield structured outputs which, in appropriate contextual configurations, display what has been called (after the lead of computer scientist Marvin Minsky) artificial intelligence. To deal with such complexities, information theory itself has been redefined as a mathematical analysis of both the transmission and manipulation of information.

According to Wiener, however, it is also the case that organisms, functioning as local anti-entropic systems, depend on Maxwell demon-like entities: "Indeed, it may well be that enzymes are metastable Maxwell demons, decreasing entropy, perhaps not by the separation between fast and slow particles but by some equivalent process." For Weiner the very essence of a living organism is that its input–output physiological functions (e.g., eating producing energy) are complemented by multiple feedback loops in which output becomes information regulating further output. In place of the static organization of nonliving matter, this creates a dynamic homeostasis through which an organism remains the same in form even while it undergoes continuous changes in the material out of which its form is constituted. This cybernetic view of biology is deepened with the discovery of DNA and the genetic code, to which information theory is again applied, and DNA is conceived as the means for transmitting an anatomy and physiology from one generation to another. The attractiveness of the information theory metaphor of a "book of life" in molecular biology has been extensively documented by Lily Kay (2000).

Cognitive psychologists were especially attracted by the idea that information theory and cybernetics could serve as the basis for a general interpretation of not only organic physiology but for much overt human behavior. In the last third of the twentieth century and with the help of computer scientists, psychologists developed increasingly sophisticated models of the mind—some attempting to replicate mental outputs by any technical means, others focused on replicating known or hypothesized brain structures and thinking processes. Independent of the extent to which the brain itself actually functions as a technical information processor, this dual research strategy proved remarkably fruitful for the investigation of human cognition and the design of advanced information processing machines. Examples in the latter case include theories of parallel information processing, neural networks, and emergent complexities.

In contrast to information theory and the technical concept of information is the concept of information in a semantic or natural linguistic sense. According to Shannon, although messages certainly have meaning, the "semantic aspects of

communication are irrelevant to the engineering problem" of efficient message transmission. Information theory, in its concern to develop a mathematical theory of technical communication, takes the linguistic meaning of any message as unproblematic and seeks to determine the most efficient way to encode and then transmit that code, which may then once again be decoded and understood. Semantic information, however, is not a two-term relation—that is, a signal being transmitted between A and B—but a three- or four-term relation: a signal being transmitted between A and B and saying something about C (possibly to D). Nevertheless, despite the difficulty of establishing an unambiguous relation between the technical and semantic notions of information (no matter what its probability, some particular signal or message may possess any number of different semantic meanings) information theory sense has had a strong tendency to influence ideas about natural language. The best background to this research program is Cherry (1978), with one representative achievement being Fred I. Dretske (1981).

Finally information theory has influenced the rational reconstruction of economic and political decision making. In this field, information theory also tends to merge with game and decision theory. After first being rejected as "bourgeois ideology," information theory and cybernetics appealed to leaders in the Soviet Union during the 1960s and 1970s as a nonideological means to reform an increasingly inefficient state planning system. Ironically enough, the same appeal has also been operative among some business and government management theorists in Europe and North America.

See also **Error Correction**

CARL MITCHAM

Further Reading

Cherry, C. *On Human Communication: A Review, a Survey, and a Criticism*, 3rd edn. MIT Press, Cambridge, MA, 1978.

Dretske, F.I. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA, 1981.

Goldstine, H.H. *The Computer: From Pascal to Von Neumann*. Princeton University Press, Princeton, NJ, 1972.

Hartley, R.V.L Transmission of information. *Bell Syst. Techn. J.*, 7, 535–563, 1928.

Kay, L.E. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford University Press, Stanford, CA, 2000.

Nyquist, H. Certain factors affecting telegraph speed. *Bell Syst. Techn. J.*, 3, 324–346, 1924.

Shannon, CE. and Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, 1949. Collates two articles: an expanded version of Weaver, W. Recent contributions to the mathematical theory of communication. *Scientific American*, July 1949; and Shannon, C.E. The mathematical theory of communication. *Bell Syst. Techn. J.* 1948.

Wiener, N. *Cybernetics: Or, Control and Communication in the Animal and the Machine*. MIT Press, Cambridge, MA, 1948.

Wiener, N. *The Human Use of Human Beings: Cybernetics and Society*. Houghton Mifflin, Boston, 1950.

Infrared Detectors

Infrared detectors rely on the change of a physical characteristic to sense illumination by infrared radiation (i.e., radiation having a wavelength longer than that of visible light). The origins of such detectors lie in the nineteenth century, although their development, variety and applications exploded during the twentieth century. William Herschel (c. 1800) employed a thermometer to detect this "radiant heat"; Macedonio Melloni, (c. 1850) invented the "thermochrose" to display spatial differences of irradiation as color patterns on a temperature-sensitive surface; and in 1882 William Abney found that photographic film could be sensitized to respond to wavelengths beyond the red end of the spectrum. Most infrared detectors, however, convert infrared radiation into an electrical signal via a variety of physical effects. Here, too, nineteenth century innovations continued in use well into the twentieth century.

Electrical photodetectors can be classed as either thermal detectors or quantum detectors. The first infrared detectors were thermal detectors: they responded to infrared radiation by the relatively indirect physical process of an increase in temperature. A thermal detector having a blackened surface is sensitive to radiation of any wavelength, a characteristic that was to become valuable to spectroscopists. The discovery of new physical principles facilitated the development of new thermal detectors. Thomas J. Seebeck reported a new "thermoelectric effect" in 1821 and then demonstrated the first "thermocouple," consisting of junctions of two metals that produced a small potential difference (voltage) when at different temperatures. In 1829 Leopoldo Nobili constructed the first "thermopile" by connecting thermocouples in series, and it was soon adapted by Melloni for radiant heat measurements—in modern parlance, for detecting infrared radiation—rather than for temperature changes produced by contact and conduction. In 1880, Samuel

P. Langley announced the "bolometer," a temperature-sensitive electrical resistance device to detect weak sources of radiant heat.

Such detectors were quickly adopted by physicists for studying optical radiation of increasingly long wavelength. Early twentieth century research in spectroscopy was largely detector-centered. This program sought to show the connections—indeed, to bridge the perceived gap—between infrared "optical" radiation and electrically generated "radio" waves. Infrared methods, for the most part, developed as an analog of visible methods while relying implicitly on electrical detectors.

During the twentieth century a variety of quantum detectors were developed and applied to a growing range of detection and measurement devices. Relying on a direct link between photons of infrared radiation and the electrical properties of the detecting material, they proved dramatically more sensitive than thermal detectors in restricted wavelength regions. As with thermal detectors, quantum detectors rely on a variety of principles. They may exhibit increased conductivity when illuminated with infrared radiation (examples of such "photoconductive" materials being pure crystals such as selenium, and compound semiconductors such as lead sulfide or lead selenide). Alternatively, quantum detectors may generate electrical current directly from infrared illumination. Examples of these "photovoltaic" detectors include semiconductor compounds such as indium antimonide or gallium arsenide.

Physical research on infrared detectors soon attracted military sponsors. Military interest centered initially on the generation and detection of invisible radiation for signaling. During the World War I, Theodore W. Case found that sulfide salts were photoconductive and developed thallous sulfide (Tl2S) cells. Supported by the U.S. Army, Case adapted these unreliable "thalofide" detectors for use as sensors in an infrared signaling device consisting of a searchlight as the source of radiation, which would be alternately blocked and uncovered to send messages (similar to smoke signals or early optical telegraphs) and a thalofide detector at the focus of a receiving mirror. With this system messages were successfully sent several miles.

During the 1930s, British infrared research focused on aircraft detection via infrared radiation as an alternative to radar; and, during World War II, relatively large-scale development programs in Germany and America generated a number of infrared-based prototypes and limited production devices.

Edgar W. Kutzscher developed the lead sulfide (PbS) photoconductive detector in Germany in 1932. This became the basis of a major wartime program during the following decade, studying the basic physics of detectors and materials, as well as production techniques and applications of infrared detection. Like the other combatants, the German military managed to deploy only limited production runs of infrared detectors and devices during World War II, for example using the radiation reflected from targets such as tanks to direct guns and developing the "lichtsprecher," or optical telephone.

In the U.S., successful developments during World War II included an infrared-guided bomb that used a bolometer as sensor, heat-sensitive phosphors for night vision "metascopes," and scanning systems used for the detection of infrared-radiating targets.

In the years following World War II, German detector technology was rapidly disseminated to British and American firms. Some of this information was recognized as having considerable military potential and was therefore classified. Infrared detectors were of great interest for locating the new jet aircraft and rockets; for their ability to be used "passively" (i.e., by measuring the radiation emitted by warm bodies rather than having to illuminate the targets with another source, as in radar); and for their increasing reliability. The potential military applications promoted intensive postwar research on the sensitivity of infrared detectors.

Whilst largely a product of military funding, these detectors gradually became available to academic spectroscopists. Improved sensitivity to infrared radiation was the postwar goal both of military designers and research scientists. The concurrent rise of infrared spectroscopy provided an impetus to improve laboratory-based detectors and led to developments such as the "Golay cell" by Marcel Golay in 1947. While this hybrid device, essentially an optically monitored pneumatic expansion cell, was a thermal detector such as the commonly used thermopile or bolometer, it was a reliable and sensitive alternative for use in spectrometers. Another new thermal detector was the thermistor bolometer, based on a blackened semiconductor (generally an oxide of a transition metal) having a narrow band-gap. Spectroscopists were also eager to discover the ultimate limitations of the newer quantum infrared detectors, and work by Peter B. Fellgett and by R. C. Jones in the early 1950s demonstrated the poor practical perfor-

mance and theoretical potential of contemporary detectors.

During this period, further developments in Germany included the thallous sulfide and lead sulfide (PbS) detectors; Americans added the lead selenide (PbSe), lead telluride (PbTe), and indium antimonide (InSb) detectors; and British workers introduced mercury–cadmium–telluride (HgCdTe) infrared detectors. The military uses found rapid application. A guided aircraft rocket (the American GAR-2) was in production by 1956, and missile guidance systems, fire control systems, bomber-defense devices, and thermal reconnaissance equipment, all employing infrared measurement devices, were available to many countries by the mid-1960s.

By the late 1970s the military technology of infrared detection was increasingly available in the commercial sphere. Further military research and development during the 1980s extended capabilities dramatically, especially for detector arrays sensitive to the mid-infrared and high background temperatures. This technology was also adapted by civilian astronomers throughout the 1980s for high-sensitivity, far-infrared use. Modern infrared detection systems fall into three distinct classes:

1. Thermal imaging devices, operating analogously to visible-light cameras
2. Low-cost single-element thermal detectors
3. Radiometric or spectrometric devices, employed for precise quantification of energy.

Detectors adapted to new, low-cost markets included the pyroelectric sensor, reliant upon the change in electrical polarization produced by the thermal expansion of a ferroelectric crystal. This is considerably more sensitive than traditional bolometers and thermopiles.

While many infrared detector principles were patented, their commercial exploitation was seldom determined in this way. Nevertheless, individual firms were able to dominate market sectors partly because of in-house manufacturing expertise, particularly for the post-World War II semiconductor detectors, or the small markets for specific detector types.

Detectors, both as single element devices and in arrays, are now increasingly packaged as components. Hence thermopiles have benefited from silicon micromachining technology, and can be manufactured by photolithography. Radiometric instruments typically are designed around such single-detector systems, although infrared arrays have been adopted for special applications, particularly astronomical observation. Such arrays, capable of precise spectroscopic, radiometric, and spatial measurement, now match thermal imagers in spatial resolution. From the Infrared Astronomy Satellite (IRAS), launched in 1983 and having a mere 62 detector elements, arrays had grown in complexity to exceed one million detector elements by 1995.

See also **Spectroscopy, Infrared**

SEAN F. JOHNSTON

Further Reading

Hudson, R.D. Jr. *Infrared System Engineering*. Wiley, New York, 1969.

Hudson, R.D. Jr. and Hudson, J.W., *Infrared Detectors*. Dowden, Hutchison & Ross, Stroudsburg, 1975.

Jha, H.R. *Infrared Technology: Applications to Electro-optics, Photonic Devices, and Sensors*. Wiley, New York, 2000.

Jamieson, J.A., McFee, R.H., Plass, G.N., Grube, R.H., and Richards, R.G. *Infrared Physics and Engineering*. McGraw-Hill, New York, 1963.

Johnston, S.F. *Fourier Transform Infrared: A Constantly Evolving Technology*. Ellis Horwood, Chichester, 1991.

Johnston, S.F. *A History of Light and Colour Measurement: Science in the Shadows*. Institute of Physics Press, Bristol, 2001.

Smith, R.E., Jones, F.E., and Chasmar, R.P. *Detection and Measurement of Infrared Radiation*. Clarendon, Oxford, 1957.

Wolfe, W.L. *Handbook of Military Infrared Technology*. Office of Naval Research, Washington D.C., 1965.

Wolfe, W.L. and Zissis, G.J. *The Infrared Handbook*. Office of Naval Research, Washington D.C., 1978.

Integrated Circuits, Design and Use

Integrated circuits (ICs) are electronic devices designed to integrate a large number of microscopic electronic components, normally connected by wires in circuits, within the same substrate material. According to the American engineer Jack S. Kilby, they are the realization of the so-called "monolithic idea": building an entire circuit out of silicon or germanium. ICs are made out of these materials because of their properties as semiconductors—materials that have a degree of electrical conductivity between that of a conductor such as metal and that of an insulator (having almost no conductivity at low temperatures). A piece of silicon containing one circuit is called a die or chip. Thus, ICs are known also as microchips. Advances in semiconductor technology in the 1960s (the miniaturization revolution) meant that the number of transistors on a single chip doubled every two years, and led to lowered microprocessor

costs and the introduction of consumer products such as handheld calculators.

In 1952, the British engineer G. W. A. Dummer of the Royal Radar Establishment in England, observed during a conference in Washington that the step following the development of transistors would be the manufacture of "electronic equipment in a solid block with no connecting wires." At the time, the American military establishment had already recognized the need to miniaturize the available electrical circuits for airplanes and missiles. Those circuits used several discrete modules such as transistors, capacitors, and resistors wired into masonite boards. Although the replacement of vacuum tubes (valves) with transistors had improved miniaturization, the circuit's manufacture was expensive and cumbersome. Thus, the military financed research programs to develop molecular electronics; that is, micromodules able to perform electronic functions on a miniaturized platform. In 1958, Kilby designed a micromodule for the firm Texas Instruments within a military-funded program. He reproduced the property of a circuit that converts direct current (DC) into alternating current (AC) into a thin wafer of germanium attached to four electrical contacts. Kilby's invention was the object of a legal controversy because in the same period, the American chemical engineers Robert Noyce and Gordon Moore, of the firm Fairchild Semiconductors, developed and patented a similar device using silicon rather than germanium. The legal controversy marked the birth of integrated circuits, and silicon eventually became the material used in the IC industry.

In the 1960s, the introduction of new types of transistors—metal-oxide semiconductors (MOS) and metal-oxide semiconductors field-effect transistors (MOSFET)—greatly improved the performance of ICs and ensured their commercial viability. In 1963, Fairchild's engineers designed ICs capable of performing resistor–transistor logic (RTL). This implied that ICs were now able to contain logic gates and perform Boolean function of the type NOT/OR/AND. The logic properties embodied in ICs made computer memory storage their main field of application. When ICs began storing computational information, their capacity was calculated in bits (binary digits). During the 1960s, ICs passed from a memory storage space capacity of 8 bits to 1024 bits. The need for greater memory space made IC design more complicated, and new techniques for computer-aided design (CAD) were thus introduced. Fairchild produced the first CAD microchip (called Micromosaic) in 1967. In the 1960s, ICs also constituted the main component in the guidance systems for spacecraft (e.g., *U.S. Freedom*, *U.S. Apollo*) and ballistic missiles' navigation systems (e.g., *U.S. Minuteman*).

In 1968, Noyce and Moore left Fairchild to establish a new company: the Integrated Electronics Corporation, or Intel. From the 1970s, their company became a leader in IC technology. In 1970, the engineers Joel Karp and Bill Regitz designed the model 1103 for Intel, the first 1024-bit dynamic random access memory (RAM) for computers. The great innovation developed by Intel, however, was the development of the first microprocessor. In 1971, Intel engineers Marcian E. Hoff, Stan Mazor, and Federico Faggin developed an IC that could be programmed with proper software and thus perform virtually any electronic function. Before 1971, each logic or arithmetic function of a computer was assigned to a specific IC, while microprocessors would work eventually as miniaturized general-purpose computers. The microprocessor was commercialized by Intel in the models 4004 (4 bits) and 8008 (8 bits). Model 4004 executed 60,000 operations in a second and deployed 2300 transistors. The introduction of microprocessors ensured an extension of IC capabilities and the development of large-scale integration (LSI), the integration of ever-increasing logic operations within the same IC.

During the 1970s, the microchip industry grew enormously, providing integrated circuits for a number of electronic devices. These commercial applications helped the IC manufacturers gain independence from military funding and to open new markets. Research and development was driven by a few companies such as Intel, mainly located in the Silicon Valley area near San Francisco. Although old electronics companies such as Texas Instruments or IBM proved strong competitors, the introduction of the model 8086 in 1978 ensured Intel's long-term success as the provider of microprocessors for computers. More generally, the development of ICs during the 1970s came to be seen as the technological basis for the so-called microelectronics revolution. According to Robert Noyce, the technological change brought by ICs would bring about a qualitative change in human capabilities. He observed that the increased number of IC components (10 in the 1960s; 1,000 in the 1970s; 100,000 and more in the 1980s) would bring about a rapid decline in the cost of given electronic functions. Furthermore, the increasing miniaturization would ease IC application in several electronic devices. Thus, ICs would become

a cheap, very reliable, and widespread technological tool in use in almost all electronic equipment.

At the end of the 1970s the American IC industry found a strong competitor in Japan, which could produce more efficient microscopic devices at a cheaper price and with more convincing design. Companies such as Nippon Electronics Corporation (NEC), Fujitsu, and Hitachi soon gained success through programs of very large scale of integration (VLSI), with ICs containing more than 1000 logic gates. Between 1977 and 1984, the American companies lost nearly 20 percent of their market to Japanese competitors. Chip development implied the definition of new designs and architectures capable of compressing functions into simpler circuits and reduced dimensions. New microprocessors also embodied the idea of reduced instruction-set computing (RISC), simplified architectures capable of operating subsets of common instructions with standard parts of microprocessors. RISC improved the speed of operation and miniaturization.

The crisis of American IC companies was overcome in the late 1980s through a new government policy that restored the traditional link between the military and the IC industry. The Pentagon funded new research for the very high-speed integrated circuit (VHSIC or vee-sick) to develop its program called the Strategic Defense Initiative (SDI) or Star Wars. Meanwhile, new materials such as gallium arsenide were introduced to replace silicon and provide faster and more reliable ICs. In the 1990s, Intel strengthened its monopoly in microprocessors for computers with the development of Pentium, the most complex processor ever built by the company after the 8086, with 3 million transistors and a speed of 100 million instructions per second.

See also **Computer Memory, Early; Integrated Circuits, Fabrication; Transistors**

SIMONE TURCHETTI

Further Reading

Augarten S. *State of the Art. A Photographic History of the Integrated Circuit*. Ticknor & Fields, New York, 1983.

Braun E. and Macdonald S. *Revolution in Miniature: The History and Impact of Semiconductor Electronics*. Cambridge University Press, Cambridge, 1982.

Jackson, T. *Inside Intel: How Andy Grove Built the World's Most Successful Chip Company*. Harper Collins, London: 1998.

Noyce, R.N. Microelectronics. *Scientific American*, 237, 3, 1977. Reprinted in *The Microelectronics Revolution*, Forester, T. Ed. Blackwell, Oxford, 1980.

Queisser H. *The Conquest of the Microchip*. Harvard University Press, Cambridge, MA, 1988.

Integrated Circuits, Fabrication

The fabrication of integrated circuits (ICs) is a complicated process that consists primarily of the transfer of a circuit design onto a piece of silicon (the silicon wafer). Using a photolithographic technique, the areas of the silicon wafer to be imprinted with electric circuitry are covered with glass plates (photomasks), irradiated with ultraviolet light, and treated with chemicals in order to shape a circuit's pattern. On the whole, IC manufacture consists of four main stages:

1. Preparation of a design
2. Preparation of photomasks and silicon wafers
3. Production
4. Testing and packaging

Preparing an IC design consists of drafting the circuit's electronic functions within the silicon board. This process has radically changed over the years due to the increasing complexity of design and the number of electronic components contained within the same IC. For example, in 1971, the Intel 4004 microprocessor was designed by just three engineers, while in the 1990s the Intel Pentium was designed by a team of 100 engineers. Moreover, the early designs were produced with traditional drafting techniques, while from the late 1970s onward the introduction of computer-aided design (CAD) techniques completely changed the design stage. Computers are used to check the design and simulate the operations of perspective ICs in order to optimize their performance. Thus, the IC drafted design can be modified up to 400 times before going into production.

Once the IC design is thoroughly checked, computers prepare the drawings of the circuit's layers. Each circuit may contain up to 15 layers and each layer defines an essential part of the circuit (gates, connections, and contacts). The drawings are then reproduced by a pattern generator in the form of sets of optical reticles that are used to create the photomasks. The reticles are up to ten times bigger than the actual chip size; hence the photomasks are produced through a photoreductive process known as step-and-repeat. The photomasks are finally cut with lithographic techniques that exploit laser beams, light, or x-rays. From the 1980s, electron-beam lithography has allowed the production of photomasks directly from computer memory, eliminating the intermediate stages.

While photomasks are manufactured, other factories are involved in the production of silicon;

the substrate material on which circuits are installed. Cylinders of raw crystalline silicon are obtained from vats of molten silicon at the temperature of 1400°C in a process that resembles the making of candles. Silicon crystals are then cut in ultra-thin wafers with a diamond saw, polished, and inspected.

At this point the manufacture of ICs begins. The method of IC's manufacture derives largely from the planar process used in the production of transistors and conceived in the late 1950s by the Swiss physicist Jean Hoerni and the American chemical engineer Robert Noyce of the American firm Fairchild Semiconductors. The planar process was a method of batch production that allowed metal connections to be evaporated onto the oxidized surface of a semiconductor. Thus the circuit areas were engraved into the substrate material. Hoerni and Noyce observed that the planar process could be easily applied only on silicon batches (rather than germanium) because in its oxidized form (silicon oxide) it allowed the introduction of insulating parts between different areas of the circuit. Thus, the isolating and conducting parts of a single component could be easily deployed and divided on the substrate material. The introduction of the planar process was one of the keys to the commercial success of ICs.

Building on the principles of the planar process, silicon wafers are covered with a layer of insulating silicon dioxide that forestalls the circuits. At this stage the photomasks come into use. The silicon wafer is exposed to ultraviolet light through the photomask. This exposure hardens some selected areas of the silicon wafer and leaves the others soft. The soft areas are then etched away with an acid bath. Recent research has allowed the definition of new dry-etching techniques without the use of corrosive acids.

The process of irradiation, treatment with chemicals, and etching is repeated at least five times. A first mask isolates the individual chips in the wafer. A second provides some areas with photoresistant material (photoresist) that defines the gates into individual chips. A third ensures the definition of the circuit's contacts by the removal of oxide parts. A fourth defines the interconnections through the introduction of metal (this is either aluminum or an alloy of aluminum and copper). Finally, a fifth mask strengthens the bonding pads through a protective glass-like substance called vapox. During some of these operations the wafers are heated to temperatures of 1000°C. The process also requires great accuracy as each photomask has to be aligned to the wafer consistently with the previous ones.

The final stage consists of testing and packaging the ICs. A first check occurs after the final overcoat on the silicon wafer passing electric current into the several ICs. Then, the wafer is cut again with the diamond saw into individual chips. At this point each chip is packaged and the IC looks like a caterpillar. The IC is tested again and the defective ones are identified and reworked.

The early stages of the IC manufacturing process are heavily automated (design, production of photomasks, silicon wafers, and ICs), while the later ones are labor-intensive (packaging and testing) and usually carried out by affiliated factories. An IC plant looks like something between a factory and a hospital. Although workers and machines fabricate chips in a highly automated context, the rooms must be clean and the workers must wear white gloves, caps, and shoe covers because even the smallest dust particle can make the circuits defective.

See also **Integrated Circuits, Design and Use; Semiconductors: Crystal Growing, Purification; Transistors**

SIMONE TURCHETTI

Further Reading

Augarten S. *State of the Art. A Photographic History of the Integrated Circuit*. Ticknor & Fields, New York, 1983.

Braun E. and Macdonald S. *Revolution in Miniature. The History and Impact of Semiconductor Electronics*. Cambridge University Press, Cambridge, 1982.

Oldham, W.G. The fabrication of microelectronic circuits. *Scientific American*, 237, 3, 1977. Reprinted in *The Microelectronics Revolution*, Forester, T., Ed. Blackwell, Oxford, 1980.

Queisser H. *The Conquest of the Microchip*. Harvard University Press, Cambridge, MA, 1988.

Intensive Care and Life Support

Intensive care is defined as the medical treatment of patients with serious reversible conditions, generally with a significant element of respiratory failure, who are judged to be capable of recovery.

The Crimean War (1853–1856) saw the start of modern nursing, pioneered by Florence Nightingale. The traditional hospital ward, with beds in rows alongside the walls was (and still is) known as a Nightingale ward. A system of progressive patient care was established in which the sickest patients were segregated and placed in the area immediately adjacent to the nursing station. In the 1940s and 1950s special areas were

developed for the care of, for example, postoperative thoracic surgical or neurosurgical patients. By the 1960s, the concept of general intensive care, or therapy, units evolved, often from anesthetist-led postoperative recovery units. The first general intensive therapy units (ITUs) were opened in the U.K. in 1966. At about the same time, coronary care units (CCUs), generally for postoperative heart surgery patients, were opened in the U.S. and the U.K. Since hospitals had been traditionally organized under a somewhat rigid hierarchical structure, the concept of multidisciplinary care with physicians, anesthetists, nurses, physiotherapists and biochemists working together was a novel one.

The impetus for these units included:

1. The increasing ability to keep patients alive using ventilators and other mechanical devices
2. Increasing understanding of the metabolic changes occurring in critical illness and the ability to monitor these changes
3. The means to treat them
4. The improvement in anesthesiology techniques that allowed complex surgery to be performed on very ill patients

Management of the systemic inflammatory response syndrome, previously referred to under the generic term "shock," gradually became better understood. Treatment of acidosis, fluid and blood replacement, and cardiovascular support with medication became the basis of intensive care. Some of the advances made possible by the isolation and treatment of patients in these specialized high technology environments define intensive care and life support.

Ventilators

The poliomyelitis epidemics of the 1930s had seen wards of patients "breathing" with the help of "iron lungs." The polio virus affected the neurological system and caused paralysis of the muscles (thus the early name of the disease, infantile paralysis), including those that make it possible to breathe. The iron lung was a bulky and cumbersome device: the patient's upper body was inside an airtight chamber in which alternate negative and positive pressures were created by a system of bellows, enabling the influx and expulsion of air from the lungs. The machine was developed by Harvard University engineer Philip Drinker and Louis Shaw in 1928. Because of iron lungs, innumerable lives were saved during the polio epidemics of the 1930s and 1940s. The supervision of these patients was almost entirely in the hands of medical students who worked a shift system. Although outdated by the development of positive pressure ventilators after World War II, 300 of these devices were amazingly still in use in the U.S. in 1985. The success of this rather primitive equipment was all the more remarkable when it is remembered that the modern, noninvasive techniques of oximetry (the measurement of blood oxygen levels) had not been invented. In 1946 a "pneumatic balance" breathing machine was described, and the Bird and Cape respirators of 1955 onward were based on this concept. Inspiratory gas flow is terminated when a preset pressure or flow is reached. Flow is generated by a mechanically driven piston, which delivers a volume of gas to the patient. Positive pressure respirators of varying degrees of complexity became the norm in hospitals although ventilation modes became much more sophisticated. By the last decade of the century it was possible to vary the pressure throughout the respiratory cycle using different waveforms and frequencies to enable patients to interact much more with the ventilator depending on their individual needs.

Cardiopulmonary Bypass

As long ago as 1812, French physician Julien Jean-Cesar LeGallois suggested that it would be theoretically possible to replace the heart with a pump. This technique required the blood to be oxygenated (receive oxygen) while it was being pumped outside the body. By the end of the nineteenth century, primitive film and bubble oxygenators had been developed. In 1916 Jay McLean, a medical student at Johns Hopkins University in Baltimore, discovered heparin, a blood component that prevents the clotting. This discovery solved the problem of blood clotting in extracorporeal circulation. In 1934, heart surgeon Michael DeBakey described the roller pump as a perfusion device. In 1944, Willem J. Kolff developed hemodialysis using a semipermeable membrane and extracorporeal blood circulation to remove impurities from the blood when the kidneys were unable to do so. John H. Gibbon Jr., performed the first "open heart" surgery using a heart–lung machine in 1953. By 1994 it was estimated that over 650,000 open-heart operations were performed worldwide annually using cardiopulmonary bypass. These machines are not without problems, which include the amount of blood required during procedures and possible damage to the patient's blood cells

with subsequent inflammatory response syndrome. To avoid these complications, there has been a trend to perform "off-pump" surgery.

Aortic Balloon Counterpulsation

During severe cardiac ischemia (lack of oxygenated blood in the tissues) or after cardiac surgery, the heart muscle can fail and lack sufficient pumping action to maintain the circulation. While awaiting definitive treatment such as a transplant or the reversal of complications as in severe disorders of cardiac rhythm, aortic balloon counterpulsation can help stabilize the patient. A balloon is positioned in the arch of the aorta, which, by precise timing of the cycle, can deflate while the left ventricle contracts (systole) forcing blood out through the arterial system, and inflate while the ventricle relaxes (diastole). This allows the maximum blood volume into circulation with each contraction and augments the coronary blood during relaxation. The device was introduced into clinical practice in 1976.

Acute Renal Failure

Modern understanding of acute renal failure (ARF) developed during the World War II years when large numbers of civilians developed renal failure following crush injuries due to collapsing buildings. As renal function failed and urinary output fell, it became apparent that too much fluid in the system was lethal. By restricting fluids to the "volume obligatoire" (i.e., just the replacement of fluid lost in perspiration and breathing), water intoxication could be avoided and renal function could recover. Following the development of dialysis techniques, ARF became somewhat more manageable.

Monitoring

Maintenance of the normal pH, or acid–base balance, of the blood is essential to health. As early as 1928, Laurence J. Henderson wrote "The condition known as acidosis is today hardly less familiar than anemia." The recognition of the importance of metabolic balance and the steady progress in resuscitation from this time culminated in the automated intensive care and critical care units of the 1990s. The patients admitted to an ICU frequently have metabolic reasons for acidosis, such as low blood pressure and lack of blood in the tissues secondary to shock, systemic infection, diabetic ketosis, or renal failure. In addition, these conditions are frequently complicated by respiratory problems, which exert a profound influence on maintenance of a normal pH because of the amount of carbon dioxide (CO_2) normally expired. The movements in pH initiated by these factors can often be in different directions. The biochemical basis for the understanding of acid–base balance was laid down between 1908 and 1920 when Henderson and K. A. Hasselbalch reported that measurement of the three variables, pH, pCO_2, and base deficit, was necessary to define acid–base status. Portable, easy-to-use blood gas machines that used blood from an arterial puncture gave rapid answers. In-dwelling arterial or venous catheters enabled frequent monitoring of other electrolytes, notably sodium, potassium, and creatinine. Continuous oximetry and simultaneous pulmonary artery pressure recording by means of disposable fiber-optic catheters had become routine by 1970. Continuous monitoring of cardiac output could be achieved by placing a Doppler probe in the esophagus. All these devices provide valuable information relevant to fluid balance and cardiovascular drug support. In addition, the status of the central nervous system can be obtained by monitoring intracranial pressure and cerebral artery blood flow.

Antibiotics

Infection, and particularly septicemia in which infection has spread through the bloodstream, either as a cause of the presenting illness or as a secondary infection, has always been a problem in ICUs. It is also among the most serious problems. Since the introduction of penicillin in 1941, innumerable increasingly potent antibiotics have been developed, only to be overused and therefore made less and less effective because of acquired resistance by microorganisms.

Conclusion

Intensive care units are not cost-effective. The technology is only manageable by dedicated one-to-one nursing by highly trained staff and the availability of an equally well-trained team of medical and other specialists. By the 1980s it had become apparent that the unrelenting stress of providing intensive care could cause psychological problems in the care givers. Research also looked at the effects on patients of sensory deprivation amid the array of technological equipment. The ability to sustain life in such an environment seemingly indefinitely had also raised questions about end-of-life care and decisions about resuscitation and life support that continued beyond the twentieth century.

See also **Dialysis**

JOHN J. HAMBLIN AND CATHERINE BROSNAN

Further Reading

Atkinson, R., Hamblin, J., and Wright, J. *Handbook of Intensive Care*. Chapman & Hall, 1981. Of historic interest only.

Fairman, J. and Lynaugh, J. *Critical Care Nursing: A History*. University of Pennsylvania Press, Philadelphia, 2000.

Galley, E. *Renal Failure: Critical Care Focus*. BMJ Books, 1999.

Hillman, K. and Bishop, G. *Clinical Intensive Care*. Cambridge University Press, Cambridge, 1996. Comprehensive up-to-date textbook.

Sasada, M and Smith, S. *Drugs in Anaesthesia and Intensive Care*. Oxford University Press, Oxford, 1997.

Sykes, K. and Young, D., Eds. *Respiratory Support in Intensive Care*. BMJ Books, 1999.

Internal Combustion Piston Engine

Born in the nineteenth century, the internal combustion piston engine did not come of age until the twentieth century. Through a long stream of refinements and expansion into a variety of sizes, shapes, and types, this device must be considered among the most influential technological developments in human history. Internal combustion piston engines, such as those used to power model aircraft, can be small enough to fit in the palm of the hand. At the other end of the spectrum, engines used to operate power generators, ships, or rail locomotives can weigh several tons and take up the space of an entire room.

As its name implies, the internal combustion engine burns fuel inside the engine itself. The most common type of internal combustion piston engine uses the four-stroke cycle (see Figure 5), also known as the ''Otto'' cycle after Nikolaus Otto, a German who patented a four-stroke engine design in 1876.

Most four-stroke engines employ poppet valves, which open and close by a camshaft to allow air and fuel to enter and exhaust to exit. Engine operation is through a sequence of four piston strokes, identified as intake, compression, power,

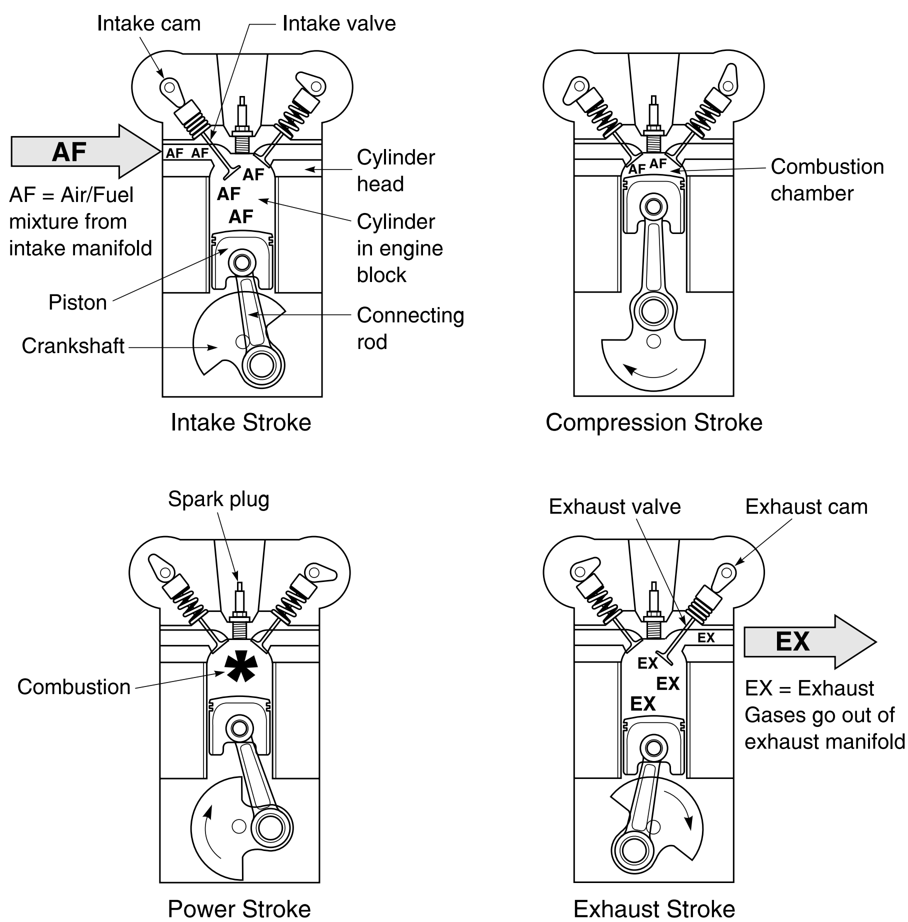


Figure 5. 4-stroke cycle engine. [*Reproduced with permission of 2 Stroke International, c 2002.*]

and exhaust. The power stroke is the point at which thermal energy is converted to kinetic energy as a result of the air/fuel mixture burning rapidly, creating high pressure in the cylinder, and pushing with great force on the head of the piston.

A two-stroke internal combustion piston engine (see Figure 6) must accomplish all of the same tasks as the four-stroke, but in half the number of strokes, identified as intake/compression stroke and power/exhaust stroke. In place of poppet valves, two-stroke engines control the flow of gases by the piston covering and uncovering ports in the cylinder, assisted by reed valves in some engines.

This arrangement gives the two-stroke engine the advantage of mechanical simplicity, and allows for smaller size and lighter weight. Early development work on modern two-stroke engine design took place in the Britain during the 1880s, prominent contributors to that work being Sir Dugald Clerk, a Scot, and Englishman James Robson.

Internal combustion piston engines are also classified based on the method of igniting the fuel. The most widely used variety is the four-stroke petrol, or gasoline, engine, in which electric spark plugs ignite the fuel, thus the designation ‘‘spark ignition.’’ Cars and trucks are the most prolific user of this type of engine and were a primary source of motivation for engine development and improvement throughout the twentieth century. Other common applications include agricultural machinery, aircraft, and military vehicles, which are discussed in detail in separate entries.

Many of the machines powered by spark-ignition gasoline engines can also operate using compression-ignition engines. This other major classification of engines is also known as the diesel engine after Rudolph Diesel, a German engineer who patented a design for a compression-ignition engine in 1893. In lieu of an electric spark, a diesel engine applies greater compression to air in the cylinder, generating enough heat to ignite the fuel as it is injected. Some diesel engines operate on the

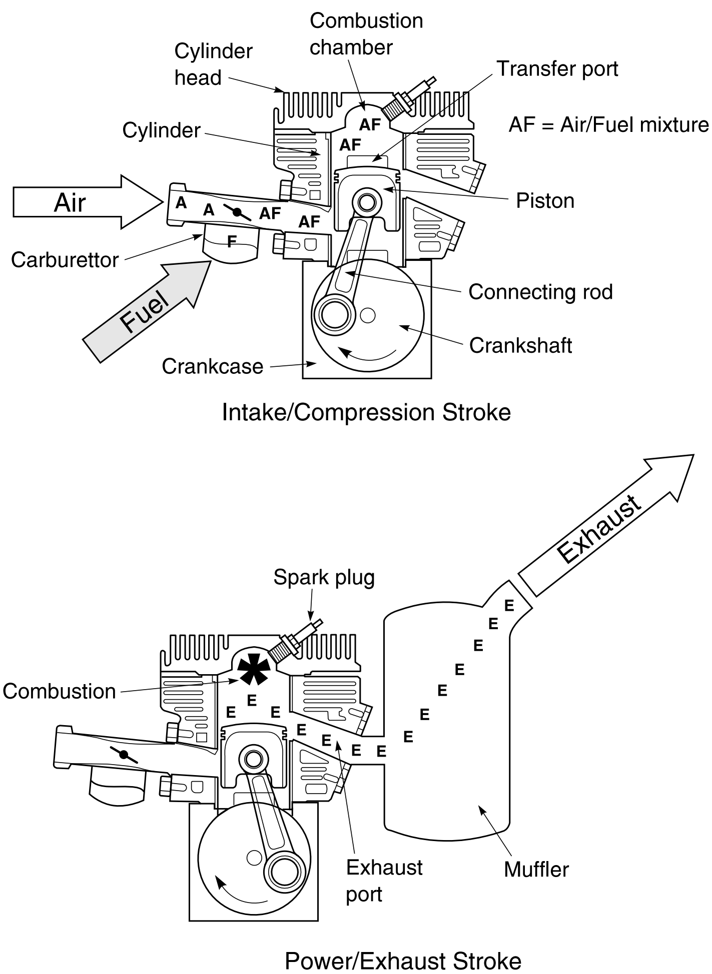


Figure 6. 2-stroke cycle engine. [*Reproduced with permission of 2 Stroke International, c 2002.*]

two-stroke cycle; however, the four-stroke cycle is more common. Diesel fuels are composed of heavier distillates of petroleum and have a greater energy density than gasoline.

Mobile uses of diesel engines include large over-the-road trucks, earthmoving and construction equipment, watercraft such as ferries and tugboats, and rail locomotives. The infusion of the internal combustion piston engine into rail transport clearly revolutionized the industry and has provided a good example of the influence of the engine as emerging technology. As the twentieth century opened, steam locomotives powered all rail transport. Their operation required that a large boiler be stoked with wood or coal, an example of an external combustion engine. A considerable time delay was involved in building up steam pressure before the locomotive could be set into motion. Frequent stops along the way were required to refill the boiler with water and the hopper car with fuel. By contrast, a diesel locomotive starts almost instantaneously, achieves greater speeds, and can travel many hundreds of miles before refueling. By the middle of the twentieth century, the steam locomotive had all but disappeared, and the diesel and diesel/electric locomotive ruled the rails.

Diesel engines are used in many stationary applications, such as water pumping stations and electrical power generators. Diesel generators are frequently used as a supplementary power source, in conjunction with electrical generating stations operating on wind or ocean tides. Backup systems for hydroelectric and even nuclear power plants often utilize diesel generators, which compared with turbine and nuclear generators are more easily transportable and operate on commonly available fuels. For these reasons, diesel generators have brought electrical power to remote regions of the world that otherwise would have none.

Diesel engines and four-stroke gasoline engines are generally larger and bulkier than two-stroke engines. Most four-strokes require an oil sump for lubrication, which means that the engine cannot be inverted or operated at differing angles of movement. Lubrication of the two-stroke engine is accomplished by mixing oil into the fuel, therefore lightweight two-stroke engines are used to power the majority of small hand-held equipment, such as chain saws and weed trimmers. Slightly larger machines such as lawnmowers, garden tillers, portable pumps and generators, jet ski-type watercraft, and outboard boat motors may be powered by either a two-stroke engine or a small four-stroke. Light motorcycles known as "dirt bikes" run on two-stroke engines, as do millions of small motor scooters common in Europe and Asia. Some compact automobiles, snowmobiles, and even small aircraft known as "ultralights" use two-stroke engines.

Many varieties of internal combustion piston engines can be fitted with minor modifications to run on natural gas, liquid petroleum gas, or hydrogen gas. As alternatives to petroleum-based fuels, renewable organic engine fuels such as alcohols (methanol and ethanol) and "biodiesel" fuels are increasing in availability and usage.

See also **Automobiles, Internal Combustion; Civil Aircraft, Propeller Driven; Farming, Mechanization; Helicopters; Motorcycles; Rail, Diesel and Diesel Electric Locomotives; Wright Flyers**

REESE E. GRIFFIN, JR.

Further Reading

Cummins, L. *Diesel's Engine*. Society of Automotive Engineers, Carnot Press, Warrendale, PA, 1993.
Cummins, L. *Internal Fire*, 3rd edn. Society of Automotive Engineers, Carnot Press, Warrendale, PA, 2000.
Griffin, Reese, Ed. *Theory Manuals on CD*. 2 Stroke International, Beaufort, SC, 2001.
Hardenberg, H.O. *The Antiquity of the Internal Combustion Engine: 1509–1688*. Society of Automotive Engineers, Warrendale, PA, 1993.
Hardenberg, H.O. *The Middle Ages of the Internal Combustion Engine: 1794–1886*. Society of Automotive Engineers, Warrendale, PA, 1999.
Heywood, JB. *Internal Combustion Engine Fundamentals*. McGraw Hill, Columbus, OH, 1988.
Heywood, J.B. *The Two Stroke Cycle Engine*. Society of Automotive Engineers, Warrendale, PA, 1999.
Obert, E.F. *Internal Combustion Engines and Air Pollution*. Addison-Wesley, Boston, MA, 1973.
Stone, R. *Introduction to Internal Combustion Engines*. Society of Automotive Engineers and MacMillan, Warrendale, PA, 1999.
Suzuki, T. *The Romance of Engines*. Society of Automotive Engineers, Warrendale, PA, 1997.

Useful Websites

2 Stroke International: http://www.2si.com
Animated Engines: http://www.keveney.comJEngines.html
DieselNet: http://www.dieselnet.com
How Stuff Works: www.howstuffworks.com
Science Museum: www.sciencemuseum.org.uk
Society of Automotive Engineers, International: www.sae.org/servlets/index

Internet

The Internet is a global computer network of networks whose origins are found in U.S. military efforts. In response to Sputnik and the emerging

space race, the Advanced Research Projects Agency (ARPA) was formed in 1958 as an agency of the Pentagon. The researchers at ARPA were given a generous mandate to develop innovative technologies such as communications.

In 1962, psychologist J.C.R. Licklider from the Massachusetts Institute of Technology's Lincoln Laboratory joined ARPA to take charge of the Information Processing Techniques Office (IPTO). In 1963 Licklider wrote a memo proposing an interactive network allowing people to communicate via computer. This project did not materialize. In 1966, Bob Taylor, then head of the IPTO, noted that he needed three different computer terminals to connect to three different machines in different locations around the nation. Taylor also recognized that universities working with IPTO needed more computing resources. Instead of the government buying machines for each university, why not share machines? Taylor revitalized Licklider's idea, securing $1 million in funding, and hired 29-year-old Larry Roberts to direct the creation of ARPAnet.

Universities were reluctant to share their precious computing resources and concerned about the processing load of a network. Wes Clark of Washington University in St. Louis, Missouri, proposed an interface message processor (IMP), a separate smaller computer for each main computer on the network that would handle the network communication. Another important idea came from Paul Baran of the RAND Corporation, who had been concerned about the vulnerability of the U.S. telephone communication system since 1960 but had yet to convince the telephone monopoly AT&T of the virtues of his ideas on distributed communications. He devised a scheme of breaking signals into blocks of information to be reassembled after reaching their destination. These blocks of information traveled through a "distributed network" where each "node," or communication point, could independently decide which path the block of information took to the next node. This allowed data to automatically flow around blockages in the network. Donald Davies at the British National Physical Laboratory (NPL) independently developed a similar concept in 1965 that he termed "packet switching," each packet being a block of data. While Baran was interested in a communications system that could continue to function during a nuclear war, ARPAnet was purely a research tool, not a command and control system.

Bolt, Beranek and Newman (BBN), a small consulting firm in Cambridge, Massachusetts, got the contract to construct ARPA's IMP in December 1968. They decided that the IMP would only handle the routing, not the transmitted data content. As an analogy, the IMP looked only at the addresses on the envelope, not at the letter inside. Faculty and graduate students at the host universities created host-to-host protocols and software to enable the computers to understand each other. They called themselves the Network Working Group (NWG) and began to share information via request for comment (RFC) papers. Steve Crocker at the University of California at Los Angeles wrote RFC Number 1, entitled "Host Software," on April 7, 1969. Because the machines did not know how to talk to each other as peers, the researchers wrote programs that fooled the computers into thinking they were talking to preexisting dumb terminals. By the summer of 1969 the NWG agreed on an overall protocol called network control protocol (NCP). RFCs are still used today to communicate about issues of Internet protocol, a versatile system for creating technical standards that allows technical excellence to dominate.

ARPAnet began with the installation of the first IMP in the fall of 1969 at UCLA, a Honeywell minicomputer built by BBN with 12 kilobyte of memory and weighing 400 kilograms, followed by three more nodes at the Stanford Research Institute (SRI), University of California at Santa Barbara, and the University of Utah. The first message transmitted between UCLA and SRI was "L," "O," "G," the first three letters of "LOGIN," then the system crashed. Initial bugs were overcome, and ARPAnet grew a node every month in 1970. In 1973 the first international nodes were added in the U.K. and Norway. More improvements and new protocols quickly followed. Electronic mail, file transfer (FTP), and remote login (Telnet) became the dominant applications. BBN invented remote diagnostics in 1970.

In 1974, Robert Kahn and Vincent Cerf proposed the first internetworking protocol, a way for datagrams (packets) to be communicated between disparate networks, and they called it an "internet." Their efforts created transmission control protocol/internet protocol (TCP/IP). In 1982, TCP/IP replaced NCP on ARPAnet. Other networks adopted TCP/IP and it became the dominant standard for all networking by the late 1990s.

In 1981 the U.S. National Science Foundation (NSF) created Computer Science Network (CSNET) to provide universities that did not have access to ARPAnet with their own network. In 1986, the NSF sponsored the NSFNET "back-

bone'' to connect five supercomputing centers. The backbone also connected ARPAnet and CSNET together, and the idea of a network of networks became firmly entrenched. The open technical architecture of the Internet allowed numerous innovations to be grafted easily onto the whole. When ARPAnet was dismantled in 1990, the Internet was thriving at universities and technology-oriented companies. The NSF backbone was dismantled in 1995 when the NSF realized that commercial entities could keep the Internet running and growing on their own, without government subsidy. Commercial network providers worked through the Commercial Internet Exchange to manage network traffic.

Other western industrialized nations also built computer networks in the mid-1970s, and gradually these also joined the Internet. The introduction in 1991 of the hypertext transport protocol (HTTP), the basis of the World Wide Web, took advantage of the modular architecture of the Internet and became widely successful. What began with four nodes as a creation of the Cold War in 1969 became a worldwide network of networks, forming a single whole. By early 2001, an estimated 120 million computers were connected to the Internet in every country of the world.

See also **Computer Networks; Electronic Communications; Packet Switching; World Wide Web**

DAVID L. FERRO AND ERIC G. SWEDIN

Further Reading

Abbate, J. *Inventing the Internet*. MIT Press, Cambridge, MA, 2000.

Hafner, K. and Lyon, M. *Where Wizards Stay Up Late: The Origins of the Internet*. Simon & Schuster, New York, 1996.

Moschovitis, C.J.P Poole, H., Schuyler, T., Senft, T.M. *History of the Internet: A Chronology, 1943 to the Present*. ABC-CLIO, Santa Barbara, CA, 1999.

Segaller, S. *Nerds 2.0.1: A Brief History of the Internet*. TV Books, New York, 1988.

Useful Websites

RFC 2555, 30 Years of RFCs: ftp://ftp.rfc-editor.org/in-notes/rfc2555.txt.

Internet Society (ISOC), histories of the Internet: http://www.isoc.org/internet/history/February 2002.

Iron and Steel Manufacture

Iron and steel manufacture is possibly the most basic industry of the twentieth century industrial economy. For much of the century, iron and steel tonnage was the most commonly used comparative measurement of a nation's economic strength.

The first step in steel manufacture is to extract iron from iron ore, then process it into steel by adding carbon and alloying elements. The predominant method of iron extraction uses a blast furnace, simply a huge shell of over 25 meters high lined with a ceramic refractory material and with holes, or taps, through which molten iron and waste products can be drawn off. The name comes from the blast of air blown into the furnace that provides oxygen for combustion. Three materials are required: iron ore, coke, and limestone. Iron ore is any of several minerals such as hematite (Fe_2O_3), limonite ($Fe_2O_3{\cdot}3H_2O$), or magnetite (Fe_3O_4), while coke, made from refining coal by heating, is the fuel for the furnace. When these materials are combined inside the furnace, the burning of coke gives off carbon monoxide, which reacts with the iron oxide as:

$$Fe_2O_3+3CO = 3CO_2+2Fe$$

The limestone serves as a scavenger to remove silica and other impurities present in the iron ore—for example, the limestone's calcium carbonate combines with silicon dioxide to form calcium silicate. The impurities float on top of the furnace melt as ''slag'' which is drawn off as waste through the ''cinder notch.'' The limestone reaction also provides some amount of extra carbon monoxide that further assists in the primary reaction. Molten iron is ''tapped'' from the furnace and cast in molds in the form of pig iron, named for the appearance of the molds as piglets gathered around a sow.

Blast furnace technology improved slowly and steadily through the twentieth century. One significant development was the pressurizing of furnaces, which began around the middle of the century and allowed more efficient combustion and increased yield. Although wrought iron, created through a process called puddling, is useful for some limited applications, pig iron is usually further strengthened by further processing into steel by the removal of impurities such as unwanted silicon, phosphorus, sulfur, and manganese and the homogenization of carbon through the solution.

The classic process for making steel is the Bessemer process developed in the nineteenth century. Air or oxygen is blown through molten pig iron in a pear-shaped, tiltable steel container called a Bessemer converter to oxidize impurities into slag. After removing the slag, the desired

alloying materials are added. The Bessemer process was highly efficient and capable of massive economic production of steel, providing steel as the basic material for the rapid expansion of the world's economy. Its disadvantage was that it required high-quality iron ore.

The basic oxygen process, or Linz–Donawitz process named for the Austrian mills that first developed it, gradually replaced the Bessemer process from the 1950s. The principle is the same; the primary difference is simply the use of pure oxygen for combustion instead of air. The basic oxygen furnace has the ability to function with lower quality ore and with up too 35 percent scrap.

Other methods of steel production were also used, including the open-hearth process. This system, which used a flat, rectangular brick hearth for oxidation of impurities, could achieve a purer metal than other processes. The use of the open-hearth process peaked in the 1950s. Because of the serious economic consequences to the industry and the often unionized laborers, controversy attended the changes in steel manufacturing at the time with intense debates over the best processes and furnaces.

The 1960s saw the introduction of the "mini-mill," in which an electric arc furnace was used to produce steel, usually of a specialized variety. The advantages of the electric furnace are tighter temperature control, no contamination from fuel, and desulfurization, while using even higher levels of scrap iron.

The production of raw steel in ingot form is only an intermediate step toward a finished product. After casting, steel is typically hot-rolled or cold-rolled into coiled sheet or strip, or into various other forms suitable for even further process: billets, skelp, slag, or bloom. There are various types of steel rolling mills, the most basic being the two-stand mill in which an ingot passes back and forth between too reversing rollers, becoming thinner and longer with each pass. A later development was the continuous rolling mill, in which stands of rollers operated in series, each at a faster rotation and smaller gap, allowing an ingot to pass through once in a single direction, eliminating the reversing and reducing cycle time. An important development in the finishing of steel was the move in the last decade of the twentieth century toward continuous casting of steel into thin coils, which reduced the number of subsequent rolling passes required.

See also **Alloys, Light and Ferrous; Nuclear Reactor Materials; Prospecting, Minerals**

FRANK WATSON

Further Reading

Chandler, H. *Metallurgy for the Non-Metallurgist*. ASM International, Materials Park, OH, 1998.
Hall, C.G.L. *Steel Phoenix: The Rise and Fall of the U.S. Steel Industry*. St. Martin's Press, New York, 1997.
Hempstead, C.A., Ed. *Cleveland Iron and Steel*. The British Steel Corporation, 1979.
McGannon, HE., Ed. *The Making, Shaping and Treating of Steel*. United States Steel Corporation, Pittsburgh, PA, 1971.
Vukmir, R.B. *The Mill*. University Press of America, Lanham, MD, 1999.

Irrigation Systems

Since the onset of human civilization, the manipulation of water through irrigation systems has allowed for the creation of agricultural bounty and the presence of ornamental landscaping, often in the most arid regions of the planet. These systems have undergone a widespread transformation during the twentieth century with the introduction of massive dams, canals, aqueducts, and new water delivery technology. In 1900 there were approximately 480,000 square kilometers of land under irrigation; by 2000 that total had surged to 2,710,000 square kilometers, with India and China as the world leaders in irrigated acreage. Globally, the agriculture industry uses about 69 percent of the available fresh water supplies, producing 40 percent of the world's food on just about 18 percent of the world's cropland. (It takes 1000 tons of water to produce 1 ton of grain.)

Traditionally flood irrigation, practiced as far back as Ancient Egypt and Sumeria, has been the standard irrigation delivery system. Flood irrigation requires relatively flat land, the construction of retaining dikes and canals, and the laborious task of breaking and filling gaps in the dikes to control the flood of water onto the field, or manually lifting the water from plot to plot.

Furrow irrigation also requires a relatively flat surface. The water from central ditches is used to fill furrows or gated pipes between rows of crops. Gated pipes are made of aluminum and have sliding, or screw-open, gates to release the water into the irrigation furrow. Siphon hoses are generally employed to deliver the water from the main ditches to the furrows. Flood and furrow irrigation are still widely employed throughout the world, but the systems require significant amounts of labor, capital, and readily available water.

With the construction of large dams, waterworks, and wells equipped with centrifugal pumps powered by electricity, diesel, gas, or steam around the globe in the twentieth century, there were also

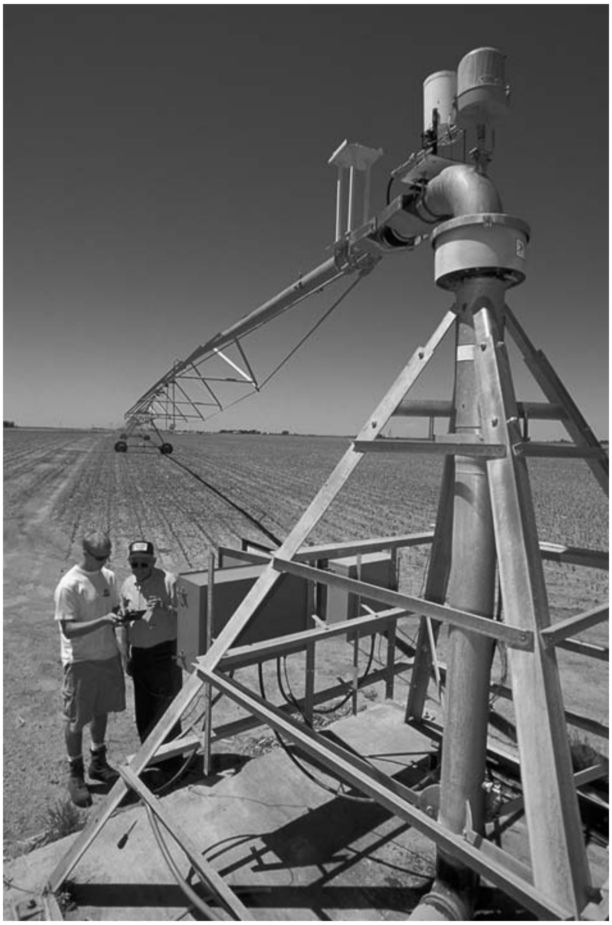


Figure 7. Agricultural Research Service scientists download data about the movement of a center-pivot irrigation system to reconstruct the amount of water and time it took to irrigate an area.
[*Photo by Scott Bauer. ARS/USDA.*]

new delivery systems for moving irrigation water to the fields. These included gated portable piping, sprinkler, and center pivot sprinkler systems, and drip irrigation systems. Massive government water projects, such as those built by the U.S. Bureau of Reclamation in the American west, led to a dramatic increase in irrigated agriculture, which quadrupled in scale from 1900 to 1990. With gas turbine pumps and deep wells, it became possible to recover, or "mine," water from underground aquifers as deep as 300 meters. Older windmills and water wheels that could provide a limited volume of water were replaced. By the year 2000, the U.S. alone had more than 500,000 such irrigation wells.

In the world's most productive farm region, the Central Valley of California where it generally rains less than 250 millimeters per year, irrigation is used on 97 percent of the acreage to produce high value crops. Irrigated by massive governmental projects that date back to the 1930s, the Central Valley is one of the most "transformed" places in the world due to irrigation. Depletion of groundwater supplies has actually led to the "sinking" of the Central Valley by several feet in some areas, although in recent years recharge basins have been refilling underground aquifers during times of water surplus. Pesticide concentration, soil salinization, and competition from urban and environmental sources have contributed to a reduction of irrigated acreage in the Central Valley since 1980.

The introduction of gated aluminum piping allowed for more efficient delivery of water to the fields where it could then be used for flood or furrow irrigation. Less water evaporates or "percolates" into the soil when aluminum irrigation pipes are used; however, the pipes are expensive and require upkeep or constant transportation from field to field. When connected to turbine pumps and improved sprinkler heads, a gated pipe system may also be used to spray a crop with water from gun-type sprinklers. Sprinkler irrigation expanded dramatically in the twentieth century in spite of problems with evaporation, salinization of fields, and high energy and equipment and maintenance costs. "Gun" sprinklers evolved from nineteenth century British firefighting technology, and by the 1870s mechanized portable, linear lateral and boom sprinkler systems had emerged, which saved labor costs. Sprinkler irrigation made it easier to control the amount of water on a crop, and it allowed irrigation on irregular surfaces.

After the 1940s in some parts of the world, such as the Great Plains of the U.S., center pivot sprinkler systems helped bring new areas under cultivation. These systems are driven by electric motors and manipulated by hydraulic systems. A giant boom with nozzles attached, the center pivot system rotates on wheels in a circle that can irrigate an entire 640-acre (2.6-square kilometer) section, thus eliminating the costly job of moving pipe. In general, sprinkler systems are expensive and result in a large amount of evaporation of water into the air, essentially wasting a resource in limited supply.

Since the 1970s drip irrigation systems have become increasingly popular throughout the world. With drip systems, underground polyvinyl chloride pipes are connected to polyethylene hoses placed above the crops. Applicators in the hoses slowly drip filtered water onto individual plants, allowing precision watering, simultaneous chemical applications, and lower water usage than other

Figure 8. Because the water level of the river varies, farmers along the Missouri River use floating pumps like this one to collect irrigation water without the pump clogging. Another concern is streambank stability, which is now more manageable thanks to an online guide developed by scientists.
[*Photo by USDA-NRCS.*]

irrigation systems. Drip irrigation systems are relatively inexpensive and easy to maintain, and they allow less evaporation than other irrigation systems.

The expansion of irrigation in the twentieth century created more productive farmland for a burgeoning world population and more forage and grain for livestock production. Water analysts predict that by 2025, 3 billion people in 48 countries will be living in conditions of water scarcity. While dam projects continue to be funded in China, Turkey, and India, the economic and ecological costs of the projects has led to a decline in development of large-scale water storage facilities.

New technologies to monitor evaporation, plant transpiration, and soil moisture levels have helped increase the efficiency of irrigation systems. The US is the world leader in irrigation technology, exporting upward of $800 million of irrigation equipment to the rest of the world each year, with the sales of drip irrigation equipment increasing 15 to 20 percent per annum in the 1990s. Golf course and landscape irrigation are also an increasing part of the irrigation technology market. Intense competition for water from cities and for environmental restoration projects might mean a reduction in irrigated agriculture in future years. At the same time, salinization of fields, infiltration of aquifers by sea water, and depleted water availability could lead to a reduction in land under irrigation worldwide.

See also **Agriculture and Food; Farming, Agricultural Methods.**

RANDAL BEEMAN

Further Reading

Fiege, M. and Cronon, W. *Irrigated Eden: The Making of an Agricultural Landscape in the American West*. Seattle, 2000.

Finkel, H.J. *Handbook of Irrigation Technology*. Boca Raton, FL, 1983.

Hurt, R.D. *Agricultural Technology in the Twentieth Century*. Manhattan, 1991.

Morgan, R.M. *Water and the Land: A History of American Irrigation*. Fairfax, 1993.

Postel, S. *Last Oasis: Facing Water Scarcity*. New York, 1992.

Reisner, M. *Cadillac Desert: The American West and Its Disappearing Water*. New York, 1986.

Useful Websites

U.S. Bureau of Reclamation Official Site: www.usbr.gov/historyborhist.htm 11.01.2003

Working Group on the History of Irrigation, Drainage and Flood Control: www.wg-hist.icidonline.org 11.01.93

(Includes references to international irrigation projects in the twentieth century.)

Irrigation Association/World Water Council: www.worldwatercouncil.org

Isotopic Analysis

Beyond the analysis of the chemical elements in a sample of matter, it is possible to determine the isotopic content of the individual chemical elements. The chemical analysis of a substance generally takes its isotopic composition to be a "standard" that represents terrestrial composition, because for most purposes the isotopic ratios are more or less fixed, allowing chemical weight to be a useful laboratory parameter for most elements. Deviations from the standard composition occur because of differences in:

1. Nuclear synthesis
2. Radioactive decay
3. Geological, biological, and artificial fractionation
4. Exposure to various sources of radiation

Applications of isotopic analysis make use of these sources of variation. Our knowledge results from analytical techniques developed during the twentieth century that allow precisions of 10^{-5} and that have led to significant improvements in the understanding of the Earth and solar system and even of archaeology.

Synthesis

Current cosmological theory holds that the chemical elements were formed in the original creation of the universe (the Big Bang), in synthesis by stars, and by the interaction of cosmic rays with interstellar matter. Within minutes after the Big Bang, all matter was composed of isotopes of hydrogen, helium, and trace amounts of lithium. Stars formed from this gas began burning hydrogen by stages to elements as heavy as iron. At this point further burning is not possible because energy is no longer released in the fusion of lighter to heavier elements, and the star undergoes a transformation dependent on its mass, initial composition, and relationship to possible companions. For stars of mass much greater than the sun, this can lead to the collapse of the iron core with the release of an enormous flux of neutrons, together with the explosion of the outer layers from nuclear reactions. The neutrons allow the synthesis of elements to uranium and beyond. Knowledge of nuclear reactions combined with theoretical models of stellar evolution allows prediction of the composition of the matter that is ejected into space. This isotopic signature varies according to the nature of the explosion, thus the solar system carries the signature of a supernova thought to have created its elements. The light elements beryllium, boron, and some additional lithium are produced by cosmic rays interacting with dust in interstellar space. The assumption that the isotopes forming the Earth were well mixed in the initial melted state cannot be relied on in detail, as special "inhomogeneities" are found within the accuracy of modern mass spectrometry.

Radioactive Decay

Most elements had radioactive isotopes in their original nucleosynthesis, but in the few million years during which the solar system formed, all but a few decayed with the daughter products joining in the composition. Some have half-lives comparable to the existence time of the solar system—most importantly uranium and thorium—which decay through a series of intermediate products ending in lead. There are about a dozen other radioactive elements scattered through the periodic table. Measuring the ratio of the parent to its daughter is the basis of determining the age when a rock solidified, but complications entailed in the history of the rock require a high level of experimental and geological skill for these determinations. These studies have attached definite ages to the epochs and periods that geologists had identified from the comparison of fossils in sedimentary rock strata. The history of the solar system during the first few million years of its formation has been established through dating based on radioactive elements, later extinct, whose daughter isotopes have been identified.

Fractionation

The isotopic composition of a chemical element can be altered through diffusion and transmutation by irradiation. Heat causes the evaporation of atoms and molecules from a sample, and the process proceeds preferentially with the lighter isotopes leaving at a greater rate than the heavy, in proportion to the ratio of the square roots of the masses. This is an important phenomenon in depleting ocean water of the isotope ^{16}O, making the ocean isotopically heavier than atmospheric oxygen. This fact is used in the principal method of determining temperatures in past ages by measuring the $^{18}O/^{16}O$ ratios at various depths in ice cores and in certain kinds of sedimentary deposits. Biochemical reactions also cause isotopic fractionation, which is

strongly expressed in the two stable isotopes of carbon and nitrogen and the four of sulfur. Differences as large as a few percent exist among the reservoirs that range from marine carbonates through a variety of plants and bacteria. Ratios of $^{34}S/^{32}S$ vary by 5 percent in coal and petroleum and 9 percent in sedimentary rocks. Diffusion is used on an industrial scale to separate isotopes with application to uranium being the best known.

Exposure to Radiation

Cosmic rays—protons with extremely high energies—can induce nuclear reactions, forming both stable and radioactive isotopes. These cosmogenic isotopes are produced on the Earth's surface and in its atmosphere and on meteorites. They figure importantly in investigations as tracers of atmospheric and geological surface processes and are important in determining time sequences in the formation of the solar system.

Star Dust

To this point, this entry has been concerned with matter in the solar system whose isotopic composition has been altered in some way. In 1987 isotopic ratios found in mineral grains of meteorites were so unusual that they forced the conclusion that the grains, which are very rare but recognizable optically, were from interstellar dust derived from the nucleosynthesis of a different star or stars than the one that formed our solar system. Measurements of the isotopic composition of these grains of "star dust" allowed the prediction of models of stellar nucleosynthesis to be compared with observation.

Archaeology

The isotopic composition of lead varies significantly according to the source from which the metal was extracted. Early metal workers introduced lead into their products either as a principal ingredient or an impurity, and measurements of their compositions offer the archaeologist clues to the origin and history of metal objects. It is a difficult subject because what is observed is the result not only of the initial metals but also of subsequent reworking by generations of craftsmen. To this end the isotopes of tin, whose primordial composition is fixed, are proving valuable. These isotopes undergo a significant fractionation with each new melting, owing to tin's volatility. Archaeologists can trace food patterns of peoples and animals with measurements of $^{13}C/^{12}C$.

See also **Mass Spectrometry**

LOUIS BROWN

Further Reading

Basu, A. and Hart, S., Eds. *Earth Processes: Reading the Isotopic Code*. American Geophysical Union, Washington D.C., 1996.

Clayton, R.N. Isotopic anomalies in the early solar system/ *Ann. Rev. of Nucl. Part. Sci.*, 28 501–522, 1981.

Cowan, J.J., Thielemann, F.-K. and Truran, J.W. Radioactive dating of the elements. *Ann. Rev. Astron. Astrophys.*, 29 447–497, 1991. Discusses the various processes of nuclear synthesis in stars.

Heaman, L. and Ludden, J.N. Applications of radiogenic isotope systems to problems in geology, in *Short Course Handbook*, 19, 1991.

Valley, J.W., and Cole, D.R., Eds. Stable isotope geochemistry. *Rev. Mineral. Geochem.*, 43, 2001.

Wetherill, G.W. Formation of the Earth. *Ann. Rev. Earth Planet. Sci.*, 18, 205–256, 1990.

Zindler, A., and Hart, S. Chemical geodynamics, in *Ann. Rev. Earth Planet. Sci.*, 14, 493–571, 1986.

J

Josephson Junction Devices

One of the most important implications of quantum physics is the existence of so-called tunneling phenomena in which elementary particles are able to cross an energy barrier on subatomic scales that it would not be possible for them to traverse were they subject to the laws of classical mechanics. In 1973 the Nobel Prize in Physics was awarded to Brian Josephson, Ivan Giaever and Leo Esaki for their work in this field. Josephson's contribution consisted of a number of important theoretical predictions made while a doctoral student at Cambridge University. His work was confirmed experimentally within a year of its publication in 1961, and practical applications were commercialized within ten years.

The device that has commonly become known as a Josephson junction consists of a thin piece of insulating material between two superconductors. The thickness of the insulator is generally of the order of 25 Å (1 Å $= 10^{-10}$ meters) or less. The electrons from the superconducting material tunnel through the nonsuperconducting insulator enabling a current to flow across the junction.

Josephson predicted two different theoretical effects for this sandwich: the first, commonly known as the DC Josephson effect, shows that a current can flow through the insulating region even when no voltage is applied across the junction; the second, the so-called AC Josephson effect, occurs when a constant voltage is applied across the junction and results in an alternating current flowing across the insulator with a high frequency in the microwave range. In the latter effect the junction acts much like an atom in which quantum mechanical transitions are taking place because electrons in superconducting solids are able to drop into a lower energy state as a result of their pairing.

When Josephson junctions are subjected to magnetic fields the current that flows across the insulator is dependent on the intensity of the field on the junction. This has important implications because when two Josephson junctions are combined in parallel as part of a superconducting loop it is possible for interference effects to occur between the junctions in an analog of the famous Young's slit experiment with light waves. In another variant only one Josephson junction is used in the superconducting loop where it acts as a parametric amplifier. Both arrangements are commonly known as superconducting quantum interference devices or SQUIDs. Modern commercially available SQUIDs have reached a high level of reliability in the manufacture of high-quality thin layers, and there has been a huge amount of research into SQUIDs. Aside from the Josephson effect they rely on another property of quantum systems, namely the quantization of magnetic flux flowing through a closed loop.

One of the most important uses of SQUIDs is as highly sensitive magnetometers, and the first SQUID magnetometer was built in 1972. It is possible to measure the miniscule magnetic fields caused by the human brain, and even the effects of thinking. As virtually all biological systems involve the movement of charged ions around the body, a corresponding magnetic field is caused. These fields are commonly of the magnitude of 10^{-13} Tesla (far smaller than the Earth's magnetic field which is of the order of 10^{-4} Tesla). A modern SQUID system for medical use may consist of up to 200 SQUIDs arranged in different axes and connected to an advanced computer

control and display system, which is capable of automatically compensating for background magnetic effects. The imaging, in real time, is interpreted as a functional map rather than a structural map. By 2002 approximately 120 hospitals in the world used low-temperature superconductor SQUID systems for cardiac (magnetocardiography) or neural imaging (magnetoencephalography), and as the technology continues to develop, this number is certain to increase.

Another interesting use of SQUIDs is in nondestructive testing of materials through measurement of magnetic fields caused by eddy currents in conducting samples. It is possible using SQUIDs to measure defects in materials nondestructively to a far greater depth than is possible using conventional techniques. As with medical imaging uses, it is common for the SQUID to be used in conjunction with a pick-up coil, which is placed adjacent to the sample. Although this technology is of great use in a laboratory setting, there are practical difficulties with using it to test materials *in situ*, such as in aerospace and civil engineering structures. These generally arise because of the requirement that the Josephson junctions must operate below a superconducting transition temperature (T_c) in order for the magnetic susceptibility to be established. For low temperature superconductors, this is of the order of 9°Kelvin (−264°C). Commercialization of portable nondestructive evaluation (NDE) systems could be as early as 2005.

Despite the use of Josephson junctions in SQUIDs, possibly the most important use of the technology will lie in future developments of the digital computer. Because Josephson junctions can operate in two different voltage states, it is possible for each junction to act as a digital switch. The switches rely on quantum transitions, and therefore it is possible for them to operate at phenomenally high speeds of a few picoseconds. When combined in a large array, this may lead to exciting developments in ultrafast digital switching and ultimately quantum computing. There are, however, a number of unfortunate practical complications in the technology as it currently stands. The most important of these is what has commonly been called the "reset" problem. Once the Josephson junction switch has been set, it is necessary to reset it by removing the background bias current, and this operation is, relatively speaking, very slow. Another issue concerns the cost of refrigeration to superconducting temperatures, which is commonly achieved using very costly liquid helium apparatus (4 to 5°Kelvin), though high-T_c materials discovered in the 1980s require cooling only to 77°Kelvin (liquid nitrogen).

Another interesting use of Josephson junctions arises when a junction is subjected to a microwave field. It is then possible for it to work in reverse so that a precise voltage is generated across the junction. When many junctions are combined it is possible to produce extremely accurate standard voltage sources, as long as the frequency of the microwaves driving the junctions is precisely known. Currently sources that generate up to 12 volts are commercially available, and it appears likely that this technology will be of increasing importance for calibration in future years. Many national metrological laboratories have adopted the Josephson effect voltage standard.

Josephson's theoretical predictions have spawned a huge field of technology since 1961 and one that seems set to become increasingly important in biomedical applications, in the testing of materials, and in the development of a quantum computer. However, perhaps more significantly, the existence and reliability of Josephson junctions provide direct confirmation of some of the most philosophically paradoxical postulates of quantum mechanics.

See also **Quantum Electronic Devices; Superconductivity, Applications**

NICHOLAS SAUNDERS

Further Reading

Barone, A. and Paternò, G. *The Physics and Applications of the Josephson Effect*. Wiley, New York, 1982.

Barone, A., Ed. *The Josephson Effect: Achievements and Trends*. World Scientific, Singapore, 1986.

Gallop, J.C. *SQUIDS: The Josephson Effects and Superconducting Electronics*. Adam Hilger, Philadelphia.

Jenks W.G, Sadeghi, S.S.H. and Wikswo, J. P. Review Article: SQUIDs for nondestructive evaluation. *J. Phys. D: Appl. Phys*. 30 293–323, 1997.

Josephson, B.D. Possible new effects in superconductive tunnelling. *Phys. Lett*., 1 251–3, 1961.

Pagano, S. and Barone, A. Josephson junctions. *Superconductor Sci. Technol*., 10 904–908, 1997.

Pizzella, V., Penna, S.D., Del Gratta, C. and Romani, G.L. Review Article: SQUID systems for biomagnetic imaging. *Superconductor Sci. Technol*., 14 R79–R114, 2001.

L

Lasers, Applications

More than 1000 different lasers (an acronym for light amplification by stimulated emission of radiation) exist and can be classified according to the type of lasing material employed, which may be a solid, liquid, gas, or semiconductor. The characteristic wavelength and power output of each laser type determines its application. The first true lasers developed in the 1960s were solid-state lasers (e.g., ruby, or neodymium: yttrium–aluminum garnet "Nd:YAG" lasers). Ruby lasers were used as early as 1961 in retinal surgery and are today used mainly in surgery and scientific research, and increasingly for micromachining. Gas lasers (helium and helium–neon being the most common, but there are also carbon dioxide lasers), developed in 1964, were soon investigated for surgical uses. Gas lasers were first used in industry in 1969 and are still heavily used in high-power applications in manufacturing for drilling, cutting, and welding. Excimer lasers use the noble gas compounds for lasing. Dye lasers, which became available in 1969, use solutions of organic dyes, such as rhodamine 6G, which can be stimulated by ultraviolet light or fast electrons. Semiconductor lasers (sometimes called diode lasers) are generally small and low power. Emitting in either the infrared or visible range, semiconductor lasers produce light based on free electrons in the conduction band, which are stimulated by an electrical current to combine with others in the valance band of the material. Covering the wavelength range used in optical fiber communications (see Lasers in Optoelectronics), semiconductor lasers are today the most important and widespread type of laser.

Lasers are employed in virtually every sector of the modern world including industry, commerce, transportation, medicine, education, science, and in many consumer devices such as CD players and laser printers. The intensity of lasers makes them ideal cutting tools since their highly focused beam cuts more accurately than machined instruments and leaves surrounding materials unaffected. Surgeons, for example, have employed carbon dioxide or argon lasers in soft tissue surgery since the early 1970s. These lasers produce infrared wavelengths of energy that are absorbed by water. Water in tissues is rapidly heated and vaporized, resulting in disintegration of the tissue. Visible wavelengths (argon ion laser) coagulate tissue. Far-ultraviolet wavelengths (higher photon energy, as produced by excimer lasers) break down molecular bonds in target tissue and "ablate" tissue without heating. Excimer lasers have been used in corneal surgery since 1984. Short pulses only affect the surface area of interest and not deeper tissues. The extremely small size of the beam, coupled with optical fibers, enables today's surgeons to conduct surgery deep inside the human body often without a single cut on the exterior. Blue lasers, developed in 1994 by Shuji Nakamura of Nichia Chemical Industries of Japan, promise even more precision than the dominant red lasers currently used and will further revolutionize surgical cutting techniques.

Commerce throughout the world has been profoundly affected by laser applications, including the widely used barcode scanning systems used in supermarkets, warehouse inventories, libraries, universities, and schools. The red "scan line" is a laser spot rapidly moving across at 30 or 40 times per second. A photo diode measures the intensity

of light reflected back from the object: since dark bars absorb light, the bar and space pattern in the barcode can be determined.

Lasers are widely employed for micromachining and automated cutting in industrial applications. Laser welding is routinely used in large industrial applications such as automotive assembly lines, providing much cheaper, better, and dramatically quicker welds than those possible with traditional techniques. Laser alloying utilizes the precise and powerful application of lasers to melt metal coatings and a portion of the underlying substrate to create surfaces that have unique and highly desirable qualities. Alloying produces the desired effect at the precise location where it is needed, meaning that less expensive materials can be utilized for the remainder of the instrument. New alloys with unique properties have been developed using this technology over the past few years with resultant new applications.

Laser diagnostic instruments have revolutionized studies across the entire range of the sciences and engineering. These include applications such as laser-induced fluorescence to measure tiny amounts of trace materials and laser Doppler anemometry, which enables fluid flow to be precisely monitored. Laser spectroscopy has revolutionized the study of very fast chemical reactions, the study of structural changes in complex molecules, and other areas of biology and chemistry. Laser photobiology and photochemistry are large and growing subdisciplines with their own conferences, journals, theories, and nomenclature.

Lasers are widely used in telecommunications, especially in fiber-optic cables. These systems employ low-powered, computer-controlled, semiconductor lasers that transmit encoded information in rapid infrared pulses. Regular light cannot perform suitably because its waves are not in parallel and therefore become too weakened over long distances resulting in an unacceptable loss of essential information. Semiconductor lasers can also "read" the pits and lands on the surface of a compact disk or DVD (see Audio Recording, Compact Disk). Because the wavelength of blue light is shorter than red light, the blue semiconductor lasers developed in the 1990s will be able to form much smaller spots on the recording layer of the disc, increasing the density of optical data storage.

The precise, unchanging nature of a laser-generated light beam makes it ideal for a wide range of applications involving measurement. Surveyors, construction personnel, oceanographers, geologists, and astronomers use laser ranging between a source and a reflector some significant distance away to measure distances or to ensure proper alignment of objects. Transit time can be used to calculate distance to an extremely high level of accuracy. Geologists routinely employ lasers, for example, to measure regional deformation of the Earth's crust to aid in understanding and predicting earthquakes. This same precision over long distances forms the core of modern laser-guided weapons, ranging from hand-held sniper rifles to long-range missiles and other "smart" weapons. Lasers are used to "paint" a target that is then precisely honed in by the weapon "tuned" to that particular wavelength. Efforts to create laser-based missile defense systems are under development in the U.S, but as of the early twenty-first century, results suggest that success is many years away (see Missiles, Defensive).

Holography is a widely employed and enjoyed aspect of the modern application of lasers. In addition to being employed for artistic and esthetic purposes, holography is used in the manufacture of optical instruments, in analyzing materials without harming them, and in storing data in extremely compact form. Two absolutely identical light beams are essential to forming a holographic image, and lasers ideally perform this function.

See also **Audio Recording, Compact Disk; Computer Memory; Lasers, Theory and Operation; Lasers in Optoelectronics; Semiconductors, Compound; Spectroscopy, Raman**

Dennis W. Cheek and Carol A. Cheek

Further Reading

Bunkin, A. and Voliak, K. *Laser Remote Sensing of the Ocean: Methods and Applications*. Wiley, New York, 2001.

Committee on Optical Science and Engineering, National Research Council. *Harnessing Light: Optical Science and Engineering for the 21st Century*. National Academy Press, Washington, D.C., 1998.

Crafer, R.C., Ed. *Laser Processing in Manufacturing*. Chapman & Hall, New York, 1993.

Diehl, R. *High-Power Diode Lasers*. Springer Verlag, New York, 2000.

Duley, W.W. *Laser Welding*. Wiley, New York, 1998.

Fotakis, C, Papazoglou, T. and Kalpouzos, C. *Optics and Lasers in Biomedicine and Culture*. Springer Verlag, New York, 2000.

Hecht, J. and Teresi, D. *Laser: Light of a Million Uses*. Dover, New York, 1998.

Nakamura, S., Pearton, S. and Fasol, G. *The Blue Laser Diode: The Complete Story*, 2nd edn. Springer Verlag, Berlin.

Puliafito, C.A. *Laser Surgery and Medicine: Principles and Practices*. Wiley, New York, 1996.

Ready, J.F. *Industrial Applications of Lasers*, 2nd edn. Academic Press, New York, 1997.

Rubahn, H.-G. *Laser Applications in Surface Science and Technology*. Wiley, New York, 1999.

Lasers in Optoelectronics

Optoelectronics, the field combining optics and electronics, is dependent on semiconductor (diode) lasers for its existence. Mass use of semiconductor lasers has emerged with the advent of CD and DVD technologies, but it is the telecommunications sector that has primarily driven the development of lasers for optoelectronic systems. Lasers are used to transmit voice, data, or video signals down fiber-optic cables.

Theodore Maiman, of Hughes Aircraft Company in the U.S., demonstrated the first laser in 1960, some four decades after the prediction of the lasing phenomenon by Einstein in 1917. Following this demonstration, in synthetic ruby crystals, it was inevitable that lasing action in other media would be investigated. It was two years later in 1962 that the Massachusetts Institute of Technology (MIT), IBM and General Electric simultaneously built the first semiconductor lasers using gallium arsenide (GaAs) as the lasing medium. Initial lasers were severely limited by heating problems and could only operate for short periods. Teams at Bell Labs, U.S. and the Ioffe Physical Institute in Russia succeeded with continuous operation at room temperature in 1970, with commercial production emerging from 1975 onwards. The development of silica optical fiber, through the 1960s and into the 1970s, has provided the complementary technology to drive and dominate the development of lasers for optoelectronics.

The basis of the semiconductor laser is a junction between *p*-type and *n*-type semiconductor materials, familiar from transistor technology. By applying a current across the junction (i.e. electrical pumping), concentrations in holes and free electrons are altered in such a way as to produce a population inversion. The injected electrons recombine with holes with the result being emission of a photon of light. Through this process of recombination in the "active region," light is emitted and, through diffusion of the carriers through the material, this light exits at the surface of the semiconductor. Natural cleavage planes of the semiconductor material form the laser end faces and define the resonant cavity within which the emission of photons of the same wavelength can build up. Once the input current reaches a certain level (usually a few milliamps, known as the "threshold current"), the optical output increases, almost linearly, with electrical input. Below the threshold, any optical emission is spontaneous, but above the threshold, the stimulated emission that defines lasing is dominant. In order to transmit information for communications purposes, the laser has to be turned on and off (modulated) and this is achieved by varying the input current by small amounts around a fixed operating current.

Devices are produced by growing semiconductor crystals, layer by layer, using epitaxial and chemical vapor deposition techniques. The wavelength of light from the laser is dependent on the band-gap energy of the semiconductor material. Thus, by selecting particular semiconductor materials and doping with various impurities it is possible to grow a range of lasers with different band-gaps, and therefore differing wavelengths of laser emission. Gallium aluminum arsenide (AlGaAs) lasers emit around 850 nanometers ($1\,\text{nm} = 10^{-9}$ meters) and were used in the first U.K. telephone field trials in 1977 at the British Post Office Research Laboratories facility at Martlesham Heath. The use of compound semiconductors from groups III and V of the periodic table continued to dominate (e.g., indium gallium arsenide phosphide (InGaAsP) which emits around 1300 nanometers). Due to low power losses (attenuation) in optical fibers at this wavelength, it remained the standard for most optical fiber systems for some time. Later generations of systems operated at 1550 nanometers (e.g., InGaAsP/InP), where lower fiber losses and the availability of optical amplifiers (erbium-doped optical fibers) at that wavelength to overcome transmission loss meant that longer spans were obtainable.

An optoelectronic communication system comprises of an electronic signal transmitted to a laser, which produces an intensity-modulated pulse stream. This modulated light signal is transmitted along optical fiber to a detector, where the optical signal is converted to electrical and amplified. When optical fibers were first proposed in the 1960s, high attenuation was a problem for long-distance communication. Removal of metal impurities achieved by Corning Glass Works (now Corning Inc.) researchers in the U.S. in 1970, kick-started a revolution in communications by fiber optic. The first commercial fiber-optic telephone systems were installed in 1977 by AT&T and GTE in Chicago and Boston respectively.

As the propagation characteristics of optical fiber (transmission losses, dispersion) were improved, so the performance characteristics of the lasers that served as the transmitters in these communication systems were refined. By adjusting

the mix of semiconductor ingredients, the peak emission wavelengths have been tuned to match transmission windows (a wavelength region that offers low optical loss) in the optical fiber; thus the greatest effort continued to be directed at improving semiconductor lasers that emit at the windows of 850, 1300, and 1550 nanometers.

With the explosion of data transmission in the 1990s, there came a need to utilize the communications infrastructure to its greatest possible capacity. This was achieved in two main ways, both made possible by developments in laser technology. First, the architecture of the semiconductor crystals evolved to enable increased modulation speeds. From 2 megabytes per second (Mb/s) in the 1970s, through 140 Mb/s in the 1980s, to 40 gigabytes per second (Gb/s) at the end of the century, there was a relentless rise in the amount of information that could be transmitted, and semiconductor laser source operating lifetimes improved greatly. Second, more optical channels were squeezed into a single fiber using dense wavelength division multiplexing (DWDM) technology. Each wavelength channel requires a separate signal-generating transmission laser. These are generally spaced at 100 or 200 gigahertz intervals across the 1530- to 1570-nanometer gain bandwidth. A wavelength multiplexer combines the separate wavelength laser signals into a single transmission fiber. Multiple wavelengths demanded increased laser stability, narrower linewidths (wavelength spread about the center wavelength) and refinement in the central wavelengths emitted. At the receiving end, a wavelength demultiplexer separates the signals into separate wavelengths of light, which are each detected at a photodiode to interface the signal into the computer or electronic network.

For a long time, semiconductor lasers emitted light parallel to the boundaries between the layered structure of the devices and were known as edge-emitters. However, surface-emitting lasers (SELs) had the advantage of being easier—and therefore more economic—to couple to other optoelectronic components. In particular, vertical cavity surface-emitting lasers (VCSELs) emerged in the late 1980s after theoretical postulation in late 1970s. Cheaper to manufacture in quantity, this laser architecture held the potential to reduce the cost of high-speed fiber optics connections, with the added benefits of low threshold current and almost circular beam symmetry (rather than elliptical), leading to improved coupling to fibers and thus efficient operation. The first VCSELs emitted at 850 nanometers, and since the 1990s lasers at 780, 859 and 980 nanometers have been commercialized into optical systems. In 1996 research started into green–blue ultraviolet devices, and at the other end of the spectrum, a VCSEL operating at 1300 nanometers was demonstrated at Sandia Laboratories in the U.S. in 2000.

While the success of lasers within telecommunication systems seems unquestioned thanks to their utility in long-distance large-capacity, point-to-point links, these lasers also find use in many other applications and are ubiquitous in the developed world. Their small physical size, low power operation, ease of modulation (via simple input current variation) and small beam size mean that these lasers are now part of our everyday world, from CDs and DVDs, to supermarket checkouts and cosmetic medicine (see Lasers, Applications).

See also **Optical Amplifiers; Optical Materials; Optoelectronics, Dense Wavelength Division Multiplexing; Semiconductors, Compound; Telecommunications**

SUSAN M. HODGSON

Further Reading

Bertolotti M. *The History of the Laser*. Institute of Physics Press, Bristol, 2004.

Bhattacharya, P. *Semiconductor Optoelectronic Devices*. Prentice Hall, Upper Saddle River, NJ, 1997.

Brinkman, W.F., Koch, T.L., Lang, D.V. and Wilt, D.P. The lasers behind the communications revolution. *Bell Lab. Tech. J.*, 5, 1, 150–167, 2000.

Dakin, J. and Brown, R. *The Handbook of Optoelectronics*. Institute of Physics Press, Bristol, 2004.

Hecht, J. *City of Light: The Story of Fiber Optics*. Oxford University Press, Oxford, 1999.

Holonyak, N. Jr. The semiconductor laser: A thirty-five year perspective. *Proceedings of the IEEE*, 85, 11, 1678–1693, 1997.

Wilson, J. and Hawkes, J. *Optoelectronics: An Introduction*. Prentice Hall, London, 1998.

Lasers, Theory and Operation

Lasers (an acronym for light amplification by stimulated emission of radiation) provide intense, focused beams of light whose unique properties enable them to be employed in a wide range of applications in the modern world (see Lasers, Applications). The key idea underlying lasers originated with Albert Einstein who published a paper in 1916 on Planck's distribution law, within which he described what happens when additional energy is introduced into an atom. Atoms have a heavy and positively charged nucleus surrounded by groups of extremely light and negatively

charged electrons. Electrons orbit the atom in a series of "fixed" levels based upon the degree of electromagnetic attraction between each single electron and the nucleus. Various orbital levels also represent different energy levels. Normally electrons remain as close to the nucleus as their energy level permits, with the consequence that an atom's overall energy level is minimized. Einstein realized that when energy is introduced to an atom; for example, through an atomic collision or through electrical stimulation, one or more electrons become excited and move to a higher energy level. This condition exists temporarily before the electron returns to its former energy level. When this decay phenomenon occurs, a photon of light is emitted. Einstein understood that since the energy transitions within the atom are always identical, the energy and the wavelength of the stimulated photon of light are also predictable; that is, a specific type of transition within an atom will yield a photon of light of a specific wavelength. Hendrick Kramers and Werner Heisenberg obtained a series of more extensive calculations of the effects of these stimulated emissions over the next decade. The first empirical evidence supporting these theoretical calculations occurred between 1926 and 1930 in a series of experiments involving electrical discharges in neon.

After World War II, a large amount of radar equipment was obtained by universities throughout the U.S. and Europe. Experiments utilizing this equipment began to lay the groundwork for the laser. By the mid-1950s, Felix Bloch, Edward Purcell, Robert V. Pound, and others had demonstrated that a significant proportion of electrons within various substances could be stimulated to raise their energy levels and then naturally decay, releasing large quantities of photons (a process known as transient population inversions). Simultaneous work on optical "pumping" techniques by Jean Brossel and Alfred Kastler demonstrated that this phenomenon could be greatly increased so that a substantially larger numbers of electrons within a substance are excited rather than unexcited. Joseph Weber described at an Electron Tube Conference in Canada in 1952 how these processes might be combined to create and maintain a large population inversion, but he was not able to produce one in the laboratory. Charles Townes at Columbia University in the U.S., working with James P. Gordon and Herbert J. Zeiger, managed such a sustained inversion process in ammonia vapor utilizing a microwave pump in 1951. In 1954 Townes formally announced the operation of an oscillator that could make this occur. Aleksandr Prokhorov and Nicolay Basov from the Lebedev Institute in Moscow published a paper in the same year about similar results using microwave spectroscopes. Both devices involved microwave amplification by stimulated emission of radiation (MASER). Townes' maser emitted at a wavelength of 1.25 centimeters and therefore did not result in visible light. In 1964 the three scientists (Townes, Prokhorov, and Basov) would all share the Nobel Prize in Physics for their work on the development of the maser-laser principle.

By 1958, Townes and Arthur Schawlow at Bell Laboratories (where Townes was also a consultant) determined that a carefully crafted optical cavity using optical mirrors would result in photons released by the pumping process being reflected back into the medium to strike other electrons. Their resultant decay from excitation would release yet further photons, the process repeating itself many times over until an intense visible beam of light would be created (all these photons would be of the same wavelength and in phase with one another). The result, light amplification by stimulated emission of radiation (LASER), would then exit the cavity through the semitransparent mirror on the one end once it had achieved sufficient intensity. The observer would see a thin beam of concentrated light. Theodore Maiman of Hughes Research Laboratories in the U.S. succeeded in turning Townes and Schawlow's ideas into reality, demonstrating in 1960 the first true laser using a ruby crystal as the laser source and a flash lamp as the source of energy "pumped" into the system. In this particular case, the two flat ends of the ruby crystal itself were coated with a reflective substance. When the flash lamp was turned on, electrons within the crystal were raised to an excited state and released photons in their decay, which were in turn reflected by the two ends of the crystal back into the medium and so on, until very quickly an intense red beam emerged from one end of the crystal. A year later, Ali Javan, William Bennett, Donald Herriott, L. F. Johnson, and K. Nassau at Bell Laboratories in the U.S., built lasers using a mixture of helium and neon gases, and a solid-state laser from neodymium (see Figure 1.) In 1962 the first semiconductor (diode) lasers using gallium arsenide were achieved, by IBM Research Laboratory in New York, General Electric in New York, and the Massachusetts Institute of Technology Lincoln Laboratory.

In 1962, Robert Hall of the General Electric Research Laboratories in the U.S. created the first infrared carbon dioxide laser, one of a new class of gas lasers whose power was unprecedented.

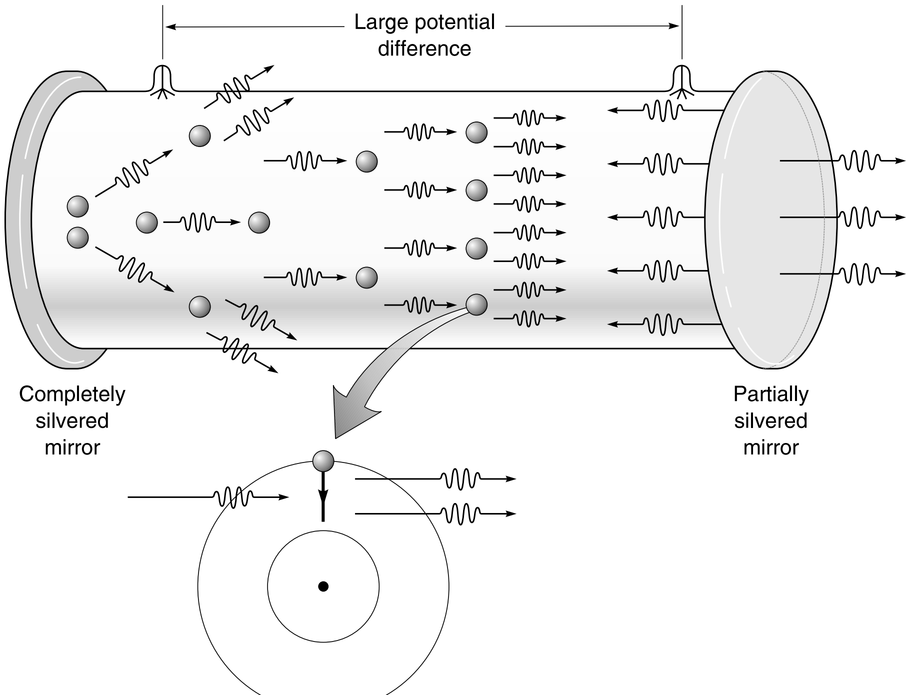


Figure 1. Schematic drawing of a helium/neon laser. [*Source: Cutnell, J.D. and Johnson, K.W. Physics, 4th edn. Wiley, New York, 1998, p. 936.*]

Creation of new types of lasers proceeded quickly over the next few years with the invention of ion lasers using mercury vapor (1963), argon (1964), and a special type of ion laser employing blue helium and cadmium metal vapor (1966). In 1966 researchers Peter Sorokin and John Lankard of the IBM Research Laboratories in the U.S. developed a liquid laser utilizing fluorescent dye that allowed investigators to "tune" certain lasers to a desired wavelength. By the mid-1970s rare gas excimer lasers such as xenon fluoride (1975) and a free-electron laser amplifier in the infrared zone (1976) were introduced. The first soft x-ray laser was successfully demonstrated by D. Matthews and colleagues at the Lawrence Livermore Laboratories in California in 1985. By the end of the twentieth century, over 1000 types of laser had been developed. Recent astronomical observations have confirmed that natural laser action also occurs in interstellar medium associated with emerging stars.

See also **Lasers, Applications; Quantum Electronic Devices**

DENNIS W. CHEEK AND CAROL A. CHEEK

Further Reading

Bertolotti, M. *Masers and Lasers: An Historical Approach*. Adam Hilger, Bristol, 1983.

Bromberg, J.L. *The Laser in America*. MIT Press, Cambridge, MA, 1991.

Committee on Optical Science and Engineering, National Research Council. *Harnessing Light: Optical Science and Engineering for the 21st Century*. National Academy Press, Washington D.C., 1998. A government-sponsored study by experts of the present and future directions of optical science and engineering, including further developments and applications of lasers.

Brown, R.G.W. and Pike, E.R. A history of optical and optoelectronic physics in the twentieth century, in *Twentieth Century Physics*, Brown, L., Pais, A. and Pippard, B., Eds. Institute of Physics Press, Bristol, 1995.

Harbison, J.P. and Nahory, R.E. *Lasers: Harnessing the Atom's Light*. W.H. Freeman, New York, 1997.

Hecht, J. *Laser Pioneers*. Academic Press, New York, 1992. Interviews with 12 scientists who were instrumental in developing masers and various kinds of lasers, including three Nobel Prize winners in Physics.

Hitz, C.B., Ewing, J.J. and Hecht, J. *Introduction to Laser Technology*, 3rd edn. Wiley, New York, 2001.

Schawlow, A.L. and Townes, C.H. Infrared and optical masers. *Phys. Rev.*, 112 1940–1949, 1958. A seminal article in the field by two of its early pioneers.

Siegman, A.E. *Lasers*, 5th revised edn. Springer Verlag, New York, 1999.

Silvast, W.T. *Laser Fundamentals*. Cambridge University Press, Cambridge, 1996.

Svelto, O. *Principles of Lasers*. 4th edn. Plenum Press, New York, 1998.

Townes, C.H. *How the Laser Happened: Adventures of a Scientist*. Oxford University Press, New York, 1999. A Nobel Prize winner's personal account of the invention of the maser and the laser and his interactions with many key scientists in the fields of applied physics and optical engineering.

Laundry Machines and Chemicals

In the nineteenth century, American home economist Catherine Beecher described laundry as the "housekeeper's hardest problem." Thousands of patents for washing machines attested to the tedium of "blue Monday," as the day commonly allocated to that task was known. The task involved hauling, boiling, and pouring hundreds of gallons of water, exposure to caustic chemicals, and rubbing, pounding, wringing, and lifting the heavy, wet linens and clothing. The automation of washing and drying machines in the twentieth century and improvements in detergents resulted in a considerable reduction in the intensity of labor required for this task. At the same time, historians of technology claim that rising standards of cleanliness and the cultural turn away from commercial laundry and domestic laundresses ensured that time spent in domestic laundering remained significant. Laundry technology never fully eliminated human effort, even in Western countries, where a person was still required to gather, sort, load, unload, fold, and iron the laundry. In less developed countries, many women continued to perform the full range of laundering tasks without machines at the end of the twentieth century.

Washing Machines

Sir John Hoskins' first British patent for a washing machine dates to 1677, and the first American patent of 1805 reflects early work on industrial machines. The British and French textile industries' use of washing machines in the bleaching process was noted in the *Grande Encyclopedie* of 1780. In fact, until the twentieth century when laundering was individualized and permanently allocated to the home, innovations in washing machines were developed principally for large commercial laundries. American James T. King's industrial cylinder washing machine of 1851 incorporated technological principles that would be applied in twentieth century automated industrial and domestic washing machines. It consisted of an outer, fixed cylinder and an inner, perforated rotating cylinder. A coal or wood fire under the outer cylinder heated soapy water that circulated through the clothes contained in the inner cylinder, which revolved by huge belts attached to an industrial steam engine. Other machines, such as Montgomery Ward's "Old Faithful" of 1927 produced well into the twentieth century, imitated the motion of the human hand rubbing against a ribbed washboard. Hand wringing to remove excess water was replaced with a hand-cranked and later electric wringer that appeared on washing machines such as the Westinghouse A4TP-2 as late as the 1960s.

By 1900, semi-automated washing machines followed two designs that would be refined and improved over the next century. In the first configuration, an exterior cylindrical shell contained a perforated inner cylindrical shell that held the clothing and rotated on a horizontal axis. In the second vertical arrangement, an outer cylinder contained an inner, perforated cylinder, or basket, in which an agitator or gyrator moved the clothing through water. The cylindrical tubs could be made of wood, copper, cast iron, and later porcelain. Later in the century, the inner shells of both designs would be configured to spin their contents at high speed in order to expel most of the water after the wash cycle was complete.

By 1939, washing machines that could agitate clothing, pipe water in and out of their tubs, and spin-dry at speed- were being marketed directly and most heavily to women. Several technological and cultural developments necessarily preceded the post-World War II mass consumption of these machines in North America and Western Europe. The electric motor, available from around the turn of the century, was initially installed in washing machines in a form that exposed users to electric shocks. It made hand cranking obsolete, although by no means in the short term. Electrical appliances in general could not become widely popular until a significant proportion of households had current delivered through an electrical network. Electrical grids by which electrical current was distributed across wide geographic areas were set up in Britain by 1926 and over a decade earlier in the U.S. Fully plumbed houses and access to piped hot water did not become the norm until after the middle of the twentieth century. Little wonder then, that a British market study of 1945 found that washing boards, dollies (hand agitators), and wringers were among the staple laundry technologies in British homes at that time.

Laundry technologies were refined at the end of the twentieth century in response to popular interest in energy conservation and environmental issues. "Smart" or fuzzy logic technologies emerged from Japan in the 1980s and were subsequently applied to major appliances, including washing machines. These technologies involved sensors that anticipated the fabric type, the turbidity (soil level), and even the hardness of the water. Total washing time, power and water level,

the amount of detergent, and the number of rinses could be regulated in the most economical manner. An example of a fuzzy rule might be: if the wash is very dirty and the dirt has a muddy quality, increase the wash time.

Chemicals

By the turn of the century, soap had become the most widespread among many chemicals used for laundering. Other products commonly used well into the century included lye, salt soda, lime, borax, and even urine and manure. Soap is made by combining animal fats or vegetable oils (glycerides) and alkalis such as sodium or potassium (once derived from ashes). Its discovery is thought to date back to Roman times. Until the twentieth century, when soap flakes and powders were introduced to the market, soaps manufactured for laundry were sold in bar or tablet form and had to be cut up or flaked by the consumer. In 1918 the British manufacturer Lever Brothers began marketing Rinso, the first granulated general laundry soap. In all these cases, the reaction between the magnesium and calcium content in hard water and soap left residual scum and was a less than satisfactory cleanser, especially in light of the rise in standards of cleanliness. Manufacturers responded by developing synthetic cleaning agents based on two-part molecules they called synthetic surfactants. The results of research at the U.S. company Proctor & Gamble were the products Dreft, followed in 1946 by Tide, which set a new standard for laundry detergents. In the final decades of the twentieth century, in response to consumer concern over the environmental impact of synthetic detergents, manufacturers developed biological and biodegradable detergents. In the former, enzymes work on biological stains. Biodegradable detergents were designed so they would not resist the biological products used to purify sewage, as had synthetic surfactants and phosphates. It was hoped that this would avoid the release of foam into rivers, which has not proven to be the case.

Among the last major innovations of the twentieth century in laundering were "detergent-less" washing machines. Korean electronics company Daewoo's Midas model and others like it were designed to use an electrolysis device that filters a number components out of tap water to produce ionized water that would clean the clothes.

See also **Detergents; Electric Motors**

JENNIFER COUSINEAU

Further Reading

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.

Davidson, C. *A Woman's Work is Never Done: A History of Housework in the British Isles 1650–1950*. Chatto & Windus, London, 1982.

Forty, A. *Objects of Desire: Design and Society from Wedgwood to IBM*. Pantheon, New York, 1986.

Gideon, S. *Mechanization Takes Command: a Contribution to Anonymous History*. W.W. Norton, London, 1948.

Hardyment, C. *From Mangle to Microwave: The Mechanization of Household Work*. Polity Press, Oxford, 1988.

Lupton, E. and Miller, J.A. *Mechanical Brides: Women and Machines From Home to Office*. Princeton Architectural Press, Princeton, 1993.

Lupton, E. and Miller, J.A. *The Bathroom The Kitchen and the Aesthetics of Waste: A Process of Elimination*. Princeton Architectural Press, New York, 1992.

Nye, D. *Electrifying America: Social Meanings of A New Technology*. MIT Press, Cambridge, MA, 1990.

Strasser, S. *Never Done: A History of American Housework*. Pantheon, New York, 1982.

Yarwood, D. *Five Hundred Years of Technology in the Home*. B.T. Batsford, London, 1983.

Useful Websites

Badami, V.V. and Chbat, N.W. *Home Appliances Get Smart*: http://www.hjtu.edu.cn/depart/xydzzxx/ec/spectrum/homeapp/app.html

Life Support, *see* **Intensive Care and Life Support**

Light Emitting Diodes

Light emitting diodes, or LEDs, are semiconductor devices that emit monochromatic light once an electric current passes through it. The color of light emitted from LEDs depends not on the color of the bulb, but on the emission's wavelength. Typically made of inorganic materials like gallium or silicon, LEDs have found frequent use as "pilot," or indicator, lights for electronic devices. Unlike incandescent light bulbs, which generate light from "heat glow," LEDs create light more efficiently and are generally more durable than traditional light sources. Despite having some evidence that Ernest Glitch produced luminescence during mid-nineteenth century experiments with crystals made from silicon and carbon, Henry Joseph Round is credited with the first published observation of electroluminescence from a semiconductor in 1907. Round, a radio engineer and inventor, created the first LED when he touched a silicon carbide (SiC) crystal with electrodes to produce a glowing yellow light. The SiC crystals, or carborundum, were widely used as abrasives in

sandpaper at the time, but the crystals were difficult to work with—and the light produced was too dim—to be of any practical use as a source of lighting. Round soon abandoned his research. Although O.V. Lossev later confirmed in 1928 that the source of the crystal's light was from "cool" luminescence, work on LEDs was largely forgotten.

Research on LEDs resumed in the 1950s, but unlike Round's use of SiC crystals, work centered on the use of gallium arsenide (GaAs) as a semiconductor. The first commercially viable LED, invented by Nick Holonyak, Jr., became available in 1962. Holonyak, who received his PhD in electrical engineering from the University of Illinois at Urbana in 1954, was the first graduate student of John Bardeen who, along with Walter Brattain, invented the transistor in 1947. After graduation, Holonyak joined Bell Labs and, in 1957, moved to General Electric (GE) laboratories in Syracuse, NY. At GE, Holonyak joined the race against other teams from GE, IBM, and the Massachusetts Institute of Technology (MIT) to develop the first infrared semiconductor laser. Robert Hall, also at GE, was the first to create a working infrared laser diode; however, Holonyak did make an interesting discovery during his research. Unlike most of his colleagues, Holonyak made his semiconductor with an alloy of gallium arsenide phosphide (GaAsP). By adding phosphorus, Holonyak increased the band gap, which reduced the wavelength of the emitted light from infrared to red. By widening the band gap, researchers discovered that LEDs could eventually produce light across the spectrum. In 1963, Holonyak returned to the University of Illinois to teach, and his former student, M. George Craford, developed a yellow LED in 1970.

From Holonyak's work, GE offered the first commercially available LED in 1962, but at a price of $260, the device proved too expensive for all but the most exclusive functions. Anticipating wider use, the Monsanto Corporation began mass-producing LEDs in 1968 and enjoyed steadily increasing sales into the next decade. In the late 1960s, electronics manufacturers began using LEDs to produce numeric displays for calculators and later wristwatches with the introduction of the Hamilton Watch Corporation's Pulsar digital watch in 1972. With time, researchers at Monsanto were eventually able to produce red, orange, yellow, and green light from LEDs, but contemporary LED technology still produced light that was too dim and consumed too much power for wider application. By the late 1970s, most manufacturers of calculators and digital watches replaced LEDs with liquid crystal displays (LCDs). Still, advances in LED technology offered much for use as indicators in devices such as telephones. The popularity of AT&T's "Princess" telephone in the 1960s, with its light-up dial, and the increasing use of multiline telephones in large corporate offices, created significant problems for the company. Existing indicators in these telephones required installation near 110-volt outlets, and the bulbs burned out often, leading to costly service calls. AT&T had much to gain from indicator lights that were both more durable and more energy efficient than current technologies. Given that LEDs had an anticipated life of over fifty years when used in telephones, the continued development of the technology offered significant cost savings for the company. Moreover, as LEDs required only about 2 volts of current they could be powered through telephone lines, which are typically around 40 volts, without affecting call quality, thereby negating the need for a separate outlet.

By the early 1980s, researchers began to develop the next generation of super high brightness LEDs. The first significant development used aluminum gallium arsenide phosphide (GaAlAsP) for LEDs that were ten times brighter than previously existing LEDs, but were only available in red and tended to have a shorter life. Research continued through the rest of the decade to achieve more colors and extra brightness without adversely effecting longevity. At the end of the 1980s, LED designers borrowed techniques from the recent developments in laser diode technology, which were gaining popular use in barcode readers, to produce visible light from indium gallium aluminum phosphide (InGaAlP) LEDs. In addition to gaining extra brightness, efficiency, and durability, the new material allowed more flexibility in changing the band gap to produce red, orange, yellow, and green LEDs. In the early 1990s, commercial applications for LEDs expanded to include traffic control devices, variable message signs, and indicator lights for the automotive industry. Also in the 1990s, researchers succeeded in creating the first blue LEDs, which made full color applications and the creation of white light from LEDs possible. By 2003, LEDs were a multibillion dollar industry. With traffic signals in the U.S. scheduled to be replaced with LEDs by 2006, the anticipated universal application of LEDs in automobiles by the end of the decade, and new uses for LEDs in medical phototherapy, this figure is sure to increase exponentially. Current designers expect

that the development of LEDs made from organic materials will further reduce the cost of production, and, as brightness and efficiency increase, some anticipate that LEDs will soon serve as a general light source that could potentially produce huge savings in annual energy costs by 2025.

See also **Computer Displays; Lighting Techniques; Semiconductors, Compound**

AARON L. ALCORN

Further Reading

Schubert, E.F. *Light-Emitting Diodes*. Cambridge University Press, New York, 2003.

Shinar, J., Ed. *Organic Light-Emitting Devices*. Springer Verlag, New York, 2003.

Zukauskas, A., Shur, M.S. and Caska, R. *Introduction to Solid-State Lighting*. Wiley, New York, 2002.

Useful websites

Hall, A. A laser in your light bulb? *Business Week Online*, 27 October 2000: http://www.businessweek.com/bwdaily/dnflash/oct2000/nf20001027_054.htm.

Martin, P. The Future Looks Bright for Solid-State Lighting. *Laser Focus World* 37, 9, September 2001: http://lfw.pennnet.com/Articles/Article_Display.cfm?Section=Articles&Subsection=Display&ARTICLE_ID =119946&KEYWORD=leds

Perry, T.S. Red hot. *IEEE Spectrum Online*, June 2003: http://www.spectrum.ieee.org/WEBONLY/publicfeature/ jun03/med.html

Rotsky, G. LEDs cast Monsanto in unfamiliar role. *Electronic Engineering Times*, 944, 10 March 1997: http://eetimes.com/anniversary/designclassics/monsanto.html

Lighting, Public and Private

At the turn of the twentieth century, lighting was in a state of flux. In technical terms, a number of emerging lighting technologies jostled for economic dominance. In social terms, changing standards of illumination began to transform cities, the workplace, and the home. In design terms, the study of illumination as a science, as an engineering profession, and as an applied art was becoming firmly established.

Overt competition between lighting methods was a relatively new phenomenon exacerbated by the spread of electric lighting during the 1880s (Edison's Lighting System and its many competitors), which rivaled the firmly entrenched gas lighting networks. During the late Victorian period, competitors marshaled science to justify the superiority of their technologies. Gas lighting, revitalized by efficient burners, incandescent mantles, and high-pressure operation, vied with the new filament electric lamps and, in larger spaces, with electric arc lighting. Between 1890 and 1910, the difficulties of incandescent lamp manufacture and the potential profits to be made from more efficient technologies motivated engineers to seek alternatives. New, more reliable, and more economical sources were continually being developed, such as the Nernst glower lamp. During this 20-year period, both innovation and technical development blossomed. The great illuminating efficiency of the firefly was much discussed, for example, and an electrochemical or luminescent analog was actively sought.

The enthusiasts who were developing lighting systems were faced with marketing, physiological, and economic questions. How were they to convince purchasers of the need for more or better lighting? How could they meaningfully compare the competing light sources in terms of brightness, color, and efficiency? How much light was needed for various tasks, and how should lighting systems best be installed and employed? Thus characteristics such as cost, intensity, and uniformity of light distribution increasingly were measured and compared. Firms devoted attention to developing more efficient (and competitive) incandescent gas mantles and electric lamp enclosures, notably the Holophane Company in New York, which produced "scientifically designed" prismatic glass globes. Yet, as more than one engineer of the period complained, the term "illumination" was more closely associated in the popular mind with medieval manuscripts or fireworks than with lighting. Indeed, perhaps the most significant alteration in the understanding of lighting during this period was the rising interest in attending to the illumination of surfaces rather than the brightness of the light source itself.

By the turn of the twentieth century, the identification of such problems was common and contributed to a growing self-awareness by practitioners of "illuminating engineering." The gas lighting industry, via local, regional and national societies, and exhibitions, was an initial nucleus for the movement. Legislators and administrators, too, had a long-standing concern with the industry owing to regulation of gas standards from the 1860s. William Preece, for example, Engineer in Chief of the Post Office in Britain, attempted to organize interest in the wider problems of lighting schemes and illumination, and the German national laboratory, the Physikalisch-Technische Reichsanstalt, undertook research for its country's gas industry during the 1890s. The PTR, in fact, effectively marshaled scientific research to measure

and regulate gas and electric lighting: the Lummer–Brodhun photometer became the standard visual device internationally. The German laboratory also undertook certification and standardization of lamps, especially after the introduction of an "illuminant tax" in 1909. Similarly, the new National Physical Laboratory in Britain (1899) and the Bureau of Standards in America (1901) made lighting standards and research an important part of their activities. The first decade of the new century saw a proliferation of interest in questions of lighting. Industrially funded lighting laboratories also appeared, such as the research arm of the National Electric Lamp Association which was founded in Cleveland in 1908. Practitioners of the lighting art in New York and London, still very much divided along the lines of their favored technologies, formed "Illuminating Engineering Societies" in 1906 and 1909, respectively; other countries followed over the next two decades. These societies, led by proselytizers Van R. Lansingh in America and Leon Gaster in Britain, sought to unite the "better educated" electrical engineers, gas engineers, physiologists, scientists, architects, and designers of lighting fixtures. An international organization, the Commission Internationale de l'Éclairage (CIE), was formed in 1913 from what had begun at a gas congress in 1900. By 1935 most European countries had similar national or regional illuminating societies contributing to the international CIE.

Such organizations promoted illumination research, standardization of lighting products, technical and public education about good lighting, and increasingly, international competition for nationally important lighting industries. A flurry of books for practitioners appeared in the years between the world wars, during which time the subject was reoriented firmly as a specialty, drawing upon electrical engineering, physics, physiology, and the psychology of perception.

An important aspect of such activities was defining of standards of illumination in public and private spaces. The interaction of safety, quality, and cost were the principal issues encouraging government involvement. In Britain, five government studies of factory and workshop lighting were carried out between 1909 and 1922. In America, the Illuminating Engineering Society published a lighting code in 1910, which led to regulations for factory lighting in five states. During the World War I, the U.S. National Defense Advisory Council issued a similar nationwide code. Purpose-designed factories of the period commonly employed large windows to maximize daylight illumination, but were inadequately supplemented by gas flames, which were often unshielded or diffused by mantles. Such dim working environments were increasingly labeled as unproductive, and an inverse correlation between illumination level and the rate of accidents was demonstrated. Moreover, workplaces frequently suffered from extremes of illumination and bright reflections. Such high contrast, or "glare," was counteracted by the employment of diffusers, shields, and reflectors over the light sources themselves and by matt finishes on walls.

Public lighting also came under increasing scrutiny. Glare in road lighting became a major concern of illuminating engineering during the 1920s and was systematically improved by careful design of the placement and elevation of road lighting fixtures and of the road surfaces themselves. During the interwar years too, signaling for vehicles, railways, roads and airports was standardized by international agreement.

The lighting of offices and homes was similarly transformed. Tables of recommended illumination had been determined empirically from the early 1890s using makeshift portable illumination photometers. Lighting levels were increasingly specified for different purposes; 136 environments ranging from coal washing to school sewing rooms to bowling alleys were specified by one source. The recommended level of lighting for offices, later correlated with working speed and accuracy, increased 25-fold over a half-century period in America: from 3 to 4 foot-candles in 1910 to 100 foot-candles in 1959. Lighting levels fell from the 1970s to 300 to 500 lux in 1993 (500 lux, or lumens per square meter, is about 47 foot-candles). This decrease in recommended lighting levels was a consequence both of higher energy costs and refinements of models of human vision, especially for lighting in conjunction with visual display units and computer usage. Other national standards and practices have risen and fallen similarly. For example, office lighting standards in Finland were 225 lux (21 foot-candles) in 1985 and 400 lux (37 foot-candles) in the Netherlands in 1991. From the 1960s, European countries began to augment such illumination standards by defining criteria for acceptable glare and color balance from lighting installations, aspects later adopted in Japan and Russia.

One reason for the rapid mid-century rise in illumination levels was the introduction of fluorescent fixtures, particularly in offices after World War II. These cylindrical sources of light have the

advantage of reducing shadows and controlling contrast in the working area and of providing high illumination levels at low electrical cost. Their disadvantages—faint hum, flicker, and a different rendition of color than incandescent lights—has limited their use in the home.

In the last decades of the twentieth century, the technological and social choices in lighting attained considerable stability both technically and socially. Newer forms of compact fluorescent lighting, despite their greater efficiency, have not significantly replaced incandescent bulbs in homes owing to higher initial cost. Low-pressure sodium lamps, on the other hand, have been adopted increasingly for street and architectural lighting owing to lower replacement and maintenance costs. As with fluorescent lighting in the 1950s, recent lighting technologies have found niche markets rather than displacing incandescents, which have now been the dominant lighting system for well over a century.

See also **Lighting Techniques**

SEAN F. JOHNSTON

Further Reading

Anon. Recommended values of illumination. *Trans. Illum. Eng. Soc. of London*, 1, 42–44, 1936.

Cady, F.E. and Dates, H.B., Eds. *Illuminating Engineering*. Wiley, New York, 1928.

Clewell, C.E. *Factory Lighting*. McGraw-Hill, New York, 1913.

Gaster, L. and Dow, J.S. *Modern Illuminants and Illuminating Engineering*. Pitman, London, 1920.

Hughes, T.P. *Networks of Power: Electrification of Western Society, 1880–1930*. MIT Press, Baltimore, 1983.

Johnston, S.F. *A History of Light and Colour Measurement: Science in the Shadows*. Institute of Physics Press, Bristol, 2001.

Trotter, A.P. *Illumination: Its Distribution and Measurement*. Macmillan, London, 1911.

Walsh, J.W.T. *Photometry*. Constable, London, 1926.

Lighting Techniques

In 1900 electric lighting in the home was a rarity. Carbon filament incandescent lamps had been around for 20 years, but few households had electricity. Arc lamps were used in streets and large buildings such as railway stations. Domestic lighting was by candle, oil and gas.

Non-Electric Lighting

Victorian gas lighting was just a burning jet of gas from a small hole, the flame little brighter than a few candles. In the 1890s the efficiency was greatly increased by the gas mantle, a fine web of rare earth oxides, which are "selective emitters," giving light only at specific wavelengths when heated. Early mantles were upright, held above the flame on a fireclay support. The "inverted" mantle, fully developed by 1905, sent light downward where it was needed. It hung from a porcelain unit that provided thermal insulation between flame and gas pipe and also deflected the combustion products away from the metal pipe.

A further advance was the high-pressure gas lamp, the subject of a practical trial in 1930. Two consecutive lengths of road in Lewisham, south London, were lit, one by high-pressure gas and the other by electric light. The local authority concluded that gas was preferable, but within a couple of years new lamps gave electricity a clear advantage.

Three other types of lighting were important where there was no electricity or gas. Most important was acetylene, known earlier but not available in any great quantity until the end of the nineteenth century. A colorless, nonpoisonous gas, it burns with a very bright flame in which colors appear as in daylight. For private houses acetylene was produced by the reaction of calcium carbide and water. Only a little gas was stored, in a gas bell floating in a tank of water. Water dropped through a control valve on to a tray of carbide. The gas passed through a pipe into the bell, which rose, operating a linkage that closed the control valve. Sometimes acetylene was used with gas mantles.

Another "home-made" gas for domestic lighting was prepared from vaporized petrol mixed with air to produce "air gas" or "petrol gas." The mixture was only explosive if the proportion of petrol was between 2 and 5 percent. Inside the drum was a rotating impeller. It forced a stream of air across gasoline stored in a separate tank. The air picked up droplets of gas.

In a simple oil lamp, the wick carries the fuel to the flame where the oil is vaporized by the heat. It is the vapor rather than the liquid that burns. Some nineteenth-century oil lamps preheated the fuel, but this only became important when incandescent mantles could exploit the increased heat of the flame. These new oil vapor lamps needed heating on starting, often using a small methylated spirit burner. In some, a small hand pump in the fuel tank helped maintain the flow. Such lamps normally had mantles and were more economical than the simple oil lamp.

Electric Lighting: Filament Lamps

The efficiency of a filament lamp increases with the temperature. In 1900 all filaments were carbon, and the practical limit was about 1600°C. At higher temperatures material evaporating from the filament blackened the glass, reducing the light output even though the lamp still functioned. The General Electric Company in America made an improved lamp, the GEM (General Electric Metallized) lamp, in 1904. During manufacture, the carbon filament was heated to 3500°C in carbon vapor. This formed a very hard outer coating and changed the electrical properties to give a positive temperature coefficient of resistance like a metal. GEM lamps operated 200°C hotter than ordinary carbon lamps and sold widely in the U.S. between 1905 and 1918. Their popularity declined as tungsten lamps became available. They were never a success in Britain or Europe where metal filaments directly succeeded the earlier carbon types.

Several metals with very high melting points were tried, including vanadium (melting point 1680°C) and niobium (1950°C). Because metals have a lower electrical resistance than carbon, metal filaments have to be longer and thinner. Furthermore, the refractory metals are brittle and difficult to draw into fine wires using conventional techniques. The initial solution was to mix powdered metal with a binder, squirt the mixture through a die, and heat the resulting thread to drive off the binder and sinter the metal particles. The first metal used successfully was osmium, a rare metal melting at about 3000°C. Osmium lamps were fragile and expensive because the long filament (typically 700 millimeters) required an elaborate "squirrel cage" support. Some lamps used an alloy of osmium and tungsten, under the name Osram, from *os*mium and wolf*ram*, the German for tungsten. These lamps were not a success, but the name was kept. In 1905 Siemens and Halske in Germany began making filament lamps with tantalum, which melts at 2996°C, and can be drawn easily. Tantalum filaments were stronger than osmium, but on alternating current circuits the metal recrystallized and became brittle.

Tungsten has the highest melting point of any metal at 3410°C. Tungsten filament lamps were first produced in Vienna by Alexander Just and Franz Hanaman, using a sintering, or heating, process. When Hugo Hirst heard of this, he set up the Osram Lamp Works at Hammersmith, which made both osmium and tungsten lamps from 1909. Manufacturing filaments by sintering was complex. After much research William Coolidge, in the General Electric research laboratory in Cleveland, Ohio, developed a process in which tungsten powder was heated and hammered, after which the metal could be drawn.

Early lamps had the bulb evacuated as completely as possible, until it was found that a little gas discouraged evaporation of the filament, increasing the life (or allowing the lamp to be run hotter for the same life). The gas, however, tended to cool the filament, reducing overall efficiency. Irving Langmuir studied the cooling of a hot wire in gas and found that although the rate of heat loss is proportional to the length of the wire, it is not much affected by the diameter. This led to winding the filament into a tight coil, which reduced the heat losses. Nitrogen-filled, coiled-filament lamps were marketed in 1913. Argon, denser and with a lower thermal conductivity, was subsequently adopted, usually with a little nitrogen (about 10 percent) because if the filament broke in pure argon an arc could form and draw a heavy current. For this reason a fuse was normally included within the lamp.

The drawn tungsten filament in a vacuum gave an efficiency of about 9 lumens per watt (1.25 watts per candle power), and the coiled filament and gas filling increased this. In 1914, Osram marketed the "Half-watt" lamp, claiming an efficiency of half a watt per candle power. This had a filament temperature of about 2800°C—some 600°C hotter than earlier tungsten lamps—and gave a whiter light.

The high working temperature necessitated changes in lamp design. Because material evaporated from the filament was carried upward by convection currents in the filling gas, the bulb was given a long neck in which evaporated material could be deposited without obscuring the light. A mica disc across the neck discouraged hot gases from overheating the cap and lamp holder.

Until the early 1920s lamps were evacuated through the end opposite the cap, leaving a characteristic, fragile "pip." Lamps are now evacuated through a tube at the cap end, protecting the sealing point within the cap. From 1934 the coiled filament was often coiled again upon itself, resulting in the modern "coiled coil" filament with an efficiency of up to fifteen lumens per watt.

To reduce glare, lamps are often made with "pearl" glass, produced by etching the surface with hydrofluoric acid. Initially, etching the glass on its inner surface resulted in a very brittle bulb, so the process was usually applied to the outside. The rough surface bulb showed finger marks and was difficult to clean. John Aldington modified the

process to give a smoother finish, making internally etched bulbs as strong as clear glass.

Attempts were made to prevent the blackening of carbon filament lamps by including chlorine in the bulb to react with the evaporated carbon and form a transparent compound. The idea was not successful with carbon lamps but was found to extend the life of tungsten lamps. Any halogen will react with tungsten vapour, and the resulting compound will break down in the region near the filament. The ultimate mechanism of failure is the same as when no halogen is present: metal evaporates from the filament leaving a thin portion which eventually overheats and then melts, but the glass is not blackened by evaporated metal. These lamps have a bulb of harder glass, with a higher softening temperature, or of fused silica (quartz). The bulb, usually much smaller than that of a conventional lamp, is enclosed in an outer bulb of ordinary glass.

Tubular halogen lamps rated at 1.5 kilowatts with a life of 2000 hours and an efficiency of 22 lumens per watt were on sale in the U.S. for floodlighting in 1960. The first halogen lamps used commercially in Europe were 200-watt lamps for airfield lighting installed at Luton Airport in the U.K. in 1961. They were also adopted for vehicle headlights where the small size of the light source was particularly advantageous. Using iodine as the halogen and quartz bulbs, they were known as quartz iodine lamps. Within a few years the quartz was superseded by hard glass, and other halogens, usually bromine, were used; the term "tungsten halogen lamps" was adopted. In the 1970s manufacturers began to use fluorine as the halogen. Tungsten hexafluoride only breaks down above 3000°C, and this temperature is achieved only at the filament itself. Consequently all the evaporated tungsten is redeposited on the filament, providing a self-healing action. The quantity of fluorine is extremely small, typically about 25 micrograms. Since the early 1990s, tungsten halogen lamps have been marketed for domestic use, both as general purpose, main-voltage lamps for higher brightness and a longer life and also as low-voltage spotlights.

Much of the energy in a filament lamp is dissipated as infrared radiation. If the heat could be returned to the filament in some way, the overall efficiency would be improved. In 1912, S.O. Hoffman obtained a patent in the U.S. for a system of prisms and mirrors designed to separate the infrared radiation and reflect it back to the filament. In 1922 he proposed the use of special coatings to achieve the effect, but it was not until the late 1970s that a potentially practical infrared reflecting layer was demonstrated. This used a so-called "silver sandwich," layers of silver and titanium dioxide arranged to reflect the heat backward while light was transmitted forward. Such lamps are ideal for display purposes because the objects being illuminated are not unduly heated. Similar layers may be used as dichroic filters to transmit some colors while reflecting others, giving a high-efficiency color filter.

Electric Lighting: Discharge Lamps

Mercury Discharge Lamps. By the end of World War I, it must have seemed that the development of electric lighting had reached its peak. Tungsten filament lamps were established as best for most purposes. Other ways of making light were being investigated, though none were as simple and convenient. Most promising were gas discharge lamps.

Several British lamp manufacturers established research laboratories. The General Electric Company (GE) opened a laboratory at Wembley in 1919, headed by Clifford Paterson. Before the outbreak of war in 1914, GE and Osram had obtained all their research support from Germany. GE's chairman, Hugo Hirst, resolved that they should never again be dependent on foreign science. Most of the lighting manufacturers cooperated because there was so much to study. The nature of the "glass" envelope, the electrodes, the filling gas, and problems of sealing the lead-in wires all needed investigation. Research in Britain concentrated on the mercury discharge lamp, although Philips in Holland favored the sodium lamp. No one could find a combination of gases in a discharge tube that would both produce white light and produce it more efficiently than a filament lamp. All early discharge lamps were more efficient and therefore cheaper to run than carbon filament lamps, but their bulky nature, very high operating voltage, and poor color were such serious disadvantages that they were never used on a large scale.

Paterson became president of the Institution of Electrical Engineers in 1930. His inaugural address was on the importance of research in electrical engineering, illustrated from two areas, heating and lighting. He thought the tungsten filament lamp had been developed to its limit, which was not quite correct, although subsequent improvements have been relatively modest. He discussed how the various gas discharge lamps might be improved, but he gave no hint of the high-pressure

mercury lamp, which his own company introduced commercially less than two years later.

The low-pressure discharge lamps of 1930 were typically glass tubes up to a meter long, which produced light of various colors for floodlighting. They were started with a high-voltage circuit employing a Tesla coil but could then run from the 230-volt mains. The color depended on the gas filling. Neon gave an orange–red light and an efficiency of up to 35 lumens per watt. Mercury gave a blue light and had a similar efficiency. In discussing "the colors most easily obtained," Paterson mentioned helium for tubes giving a yellow light. He also said that sodium vapor offered the possibility of a lamp giving up to 50 lumens per watt. At that time the chemical problems posed by the high reactivity of hot sodium vapor had not been solved.

The practical high-pressure mercury vapor lamp dates from July 1932, when GE installed them on the road outside their laboratories. The first were rated at 400 watts and had an arc 160 millimeters long in an aluminosilicate glass discharge tube 34 millimeters in diameter. The discharge tube contained a little argon and sufficient mercury to give a vapor pressure of 1 atmosphere at the operating temperature of 600°C. To reduce heat losses, the discharge tube was mounted in an evacuated outer bulb, giving an efficiency of about 36 lumens per watt and a life of 1200 hours.

Detailed improvements between 1932 and 1939 led to a range of lamps rated between 150 and 750 watts, with a life of up to 7000 hours and an efficiency of up to 42 lumens per watt. The arc tube was often made of quartz, which required new ways of sealing in the connecting wires. Platinum wires had been used in the earliest filament lamps because platinum and soda glass have almost the same coefficient of expansion, but a platinum wire sealed through quartz shrinks on cooling while the quartz maintains its dimensions. Similarly the dumet wire that replaced platinum in ordinary tungsten filament lamps cannot be used with quartz. Molybdenum was eventually adopted: a thin molybdenum wire could be sealed satisfactorily through quartz. A thicker wire did not make a good seal until Aldington found that if part of the wire were rolled flat, a good seal could be obtained around the flattened portion.

Much effort was devoted to improving the color of the light. An early approach was the "blended" lamp, a mercury lamp incorporating a tungsten filament run at a low temperature. The filament acted as the ballast resistance to control the current in the circuit and contributed red light to improve the overall color. Other lamps incorporated the vapor of metals such as zinc and cadmium to improve the color spectrum of the output. These "high brightness" lamps concentrated the discharge in a small volume between massive tungsten electrodes. The arc length of less than 1 centimeter was ideal for projectors. In 1947 they were introduced in film studios where previously the carbon arc had been the only suitable light.

Even more compact mercury discharge lamps with the arc in a narrow, water-cooled, quartz tube were developed in World War II as a substitute for the carbon arc in searchlights, but the complication of water cooling made them impractical.

Sodium Discharge Lamps. The simple, low-pressure, sodium lamp is the most efficient way known of producing light. Made in sizes from 10 to 180 watts, with efficiencies ranging from 100 to 180 lumens per watt, it has proved popular for street lighting, where the operating cost is a dominant consideration, even though the yellow light is unattractive. The spectrum of a low-pressure sodium lamp contains two lines, at 589.0 and 589.6 nanometers, close together in the yellow part of the spectrum so that for practical purposes the light is monochromatic, and colors cannot be seen by it.

The sodium lamp works in a similar way to the mercury lamp. Sodium is solid at ordinary temperatures, so the lamps contain a little neon in which a discharge starts with a red glow when the lamp is first switched on. As the lamp warms, which takes several minutes, the sodium vaporizes and takes over the arc. The neon ceases to play any part. To reduce heat losses and improve efficiency, sodium lamps are enclosed in a vacuum jacket, which in early lamps was detachable and reusable.

Sodium is highly chemically reactive. Glasses containing silica are attacked and blackened by hot sodium. Silica-free aluminoborate glass will withstand hot sodium, but it is attacked by atmospheric moisture and difficult to work. Low-pressure sodium lamps have an envelope of soda-lime glass coated on the inside with a very thin layer of aluminoborate glass. The electrodes are of oxide-coated tungsten similar to those in the high-pressure mercury lamp.

The high-pressure sodium lamp, introduced commercially in the 1970s, is nearly as efficient as the low-pressure lamp and gives a wide-spectrum, "salmon pink" light in which most colors are readily seen. The arc tube is made of alumina (Al_2O_3), a translucent ceramic, which is ground to a fine powder, pressed, and sintered in the required

tubular shape. Alumina melts at 2050°C and is one of the most stable substances known, occurring naturally as sapphire and ruby; it is unaffected by hot sodium. The alumina tube can be cut to length but cannot otherwise be worked. In one construction method, the lead-in wires supporting the electrodes are sealed through an end cap, also made of alumina, and sealed to the tube with a sodium-resistant glass. The electrodes are oxide-coated tungsten connected by a niobium lead-in wire. The filling gas is sodium with a little mercury and either argon or xenon to assist in starting.

Electric Lighting: Fluorescent Lamps

The electric light most commonly found in shops and offices, and increasingly in the home, is the fluorescent tube, available in many shapes, sizes, and colors. These also exploit an electric discharge in mercury but use the ultraviolet light that is produced. A phosphor coating on the inside of the glass tube converts this ultraviolet into visible light. The glass is opaque to ultraviolet light so none escapes.

Although efficient, early fluorescent lamps were not popular because of poor color rendering and tendency to flicker, especially when old. They needed special control equipment, making their initial cost greater than that of filament lamps, a further disadvantage. The best modern fluorescents have largely overcome these problems and, with their low operating costs, are rapidly increasing their market share.

By 1930 it was known that a mixture of mercury vapor and another gas could convert as much as 60 percent of the electrical energy in a discharge into radiation at a single ultraviolet wavelength, 253.7 nanometers. Cold cathode fluorescent tubes, operating at several thousand volts, were introduced for lighting large spaces in 1938, quickly followed by the familiar hot cathode tube operating at mains voltage. Commercial exploitation of fluorescent lighting was delayed by the onset of World War II. Although it was used in factories, it was not commonly seen by the general public until the late 1940s.

Nearly all fluorescent tubes were made with a diameter of 38 millimeters until the early 1970s because that gave the greatest efficiency. The most widely used tube was the 40-watt, 1200-millimeter-long, 38-millimeter-diameter lamp. Continuing improvements in phosphors and electrodes steadily improved efficiency and life. Most of the mineral-based phosphors used in the first fluorescent tubes were sulfides such as cadmium sulfide and zinc sulfide. From the late 1940s they were replaced by halophosphates, a British discovery of 1942. From about 1970 alkaline-earth silicate phosphors have been developed. These are more efficient than the halophosphates, but, whereas the halophosphates give a broad spectrum, the newer materials usually give light with a single line spectrum. In most fluorescent lamps a mixture of fluorescent materials is used to obtain the desired color and efficiency. Most modern fluorescent lamps employ several narrow-band phosphors of distinct colors as may be seen by viewing their light through a prism.

From 1975 the "slimline" fluorescent tubes, 26 millimeters in diameter, gradually superseded the 38-millimeter ones. The 26-millimeter tubes have krypton rather than argon for the inert gas filling and use 10 percent less electricity. Ever narrower lamps continue to be developed, using more efficient phosphors and operating at higher power densities than earlier phosphors.

The compact fluorescent lamp, launched by Philips in 1980, was designed to be a direct replacement of the incandescent lamp. The control circuitry was contained within its cap. It had four times the efficiency and five times the life of a filament lamp. A variant introduced in 1982 by Thorn Lighting in Britain was their two-dimensional lamp, which was plugged into the control gear, a separate, reusable unit that was then placed into a conventional lamp. Early compact fluorescents were heavy because the current was controlled by an iron-cored inductance, which has now been replaced by the electronic ballast. Electronic control has other advantages: the lamp starts without flicker after a half-second delay, during which the electrodes are prewarmed before the full voltage is applied, and then runs at a more efficient frequency of about 25,000 Hertz (compared with the mains frequency of 50 Hertz). Although a direct descendant of the long fluorescent tube, the compact fluorescent is virtually a new lamp and has been marketed vigorously for its "energy saving" qualities.

Electric Lighting: Other Lamps

In addition to mercury or sodium, many other elements can provide the medium for a light-giving discharge. The metal halide lamp (not to be confused with the tungsten halogen lamp) introduced in the 1960s was a logical development from high-pressure mercury lamps, containing additional metal vapors to improve the color. Several other metals are potentially suitable, but in the

high-temperature and high-pressure conditions within the arc tube they would react chemically either with the silica of the tube or with the electrodes and lead-in wires. The successful metal halide lamp depends on the discovery that such chemical attack can be avoided by including the metals as their halides, usually the iodide. Metals used include dysprosium, gallium, indium, lithium, scandium, sodium, thallium and thorium. When the lamp is cold, the metal in its halide form is condensed on the silica tube. When switched on, it starts as a high-pressure mercury lamp, but the metal halides quickly vaporize, and in the intense heat they dissociate. The metal vapor then produces its own characteristic emission in the arc, but any metal atoms that diffuse toward the silica wall or the electrodes, which are cool compared with the arc, recombine with the halogen before they can cause damage by chemical attack.

Metal halide lamps are similar in construction to high-pressure mercury lamps, though usually smaller. They require similar starting and control circuits, and are widely used in large lighting schemes where bright light of good color is required, such as sports fields or arenas and television studios. Their efficiency (typically 70 to 80 lumens per watt) is better than the high-pressure mercury lamp (52 lumens per watt) though not as high as the high-pressure sodium lamp (115 lumens per watt), but their color is better than either.

The ''induction'' lamp, also introduced in the 1990s, is a development of the fluorescent lamp. Conventional fluorescent lamps eventually fail either because the electrodes wear out or because the seal between the glass and the lead-in wires fails. The induction lamp has no electrodes and no lead-in wires; the discharge is created by a very high-frequency electromagnetic field outside the bulb.

Completely new at the time was the ''sulfur'' lamp, an induction lamp invented in 1990 by Fusion Lighting of Maryland. A glass bulb the size of a golf ball contains sulfur and argon. When energized from a high frequency source, it gives a bright, very white light whose spectral composition is similar to sunlight. It produces very little ultraviolet. Each sulfur lamp absorbs 5,900 watts but gives as much light as 25,000 watts of incandescent lighting. The high frequency supply comes from a magnetron. The lamp itself is environmentally friendly because it contains no mercury or any other toxic substances.

Some semiconductor devices emit light. These light emitting diodes, or LEDs, are often used for small signs and indicator lamps. They are very efficient light sources but were long limited to low powers and a restricted color range. At the start of the century, bright green LEDs became available, and LEDs are appearing in traffic lights. LEDs seem likely to be an important light source in the twenty-first century.

See also **Lighting, Public and Private; Light Emitting Diodes**

BRIAN BOWERS

Further Reading

Cady, F.E. and Dates, H.B., Eds. *Illuminating Engineering*. Wiley, New York, 1928.

Gaster, L. and Dow, J.S. *Modern Illuminants and Illuminating Engineering*. Pitman, London, 1920.

Hughes, T.P. *Networks of Power: Electrification of Western Society, 1880–1930*. MIT Press, Boston, 1983.

Johnston, S.F. *A History of Light and Colour Measurement: Science in the Shadows*. Institute of Physics Press, Bristol, 2001.

Liquid Crystals

The term ''liquid crystals'' contains a contradiction in terms in that crystals are solid, not liquid. But a number of materials can be in a phase that is in between the solid and the liquid phase, and the term liquid crystals refers to this phase, called the mesomorphic phase, rather than to a specific class of materials. Meso refers to middle, as this state in is the middle between solid and liquid. On the other hand, liquid crystals combine characteristics of both the solid crystalline phase and the liquid phase. They have the mobility of a liquid, but also the anisotropy (anisotropy means that the properties mentioned depend on the direction in which they are measured) in optical and electrical properties that otherwise is only found in crystals. This makes them suitable for a number of applications.

The mesomorphic phase was discovered in 1888 by an Austrian botanist named F. Reinitzer. He found that when cholesterol in plants was being melted, it passed not one but two transition points. The first one marked the transition from solid into a hazy liquid at a certain temperature (the melting point), and the second one marked the transition from a hazy into a transparent liquid at a higher temperature (the clearing point). Moreover he observed the phenomenon of birefringe: the liquid crystal differentiates in diffraction index between light with different polarization directions. In 1922 G. Friedel developed a classification that is still used today and which is discussed below. D. Vorländer contributed to the interest in liquid

crystals by synthesizing numerous organic compounds that showed the mesomorphic phase properties. In 1933 the Royal Society held a conference on liquid crystals. It was so successful in suggesting that all possible knowledge about liquid crystals had been acquired that scientific interest in liquid crystals decreased rapidly. In 1957 there was a revival following the publication of a review article on liquid crystals by Glenn Brown, an American chemist, which nearly coincided with the finding of Westinghouse researchers that liquid crystals could be used to make temperature sensors. In the 1960s, research into liquid crystals greatly intensified because a number of other applications were envisioned, and again several new compounds were developed. The main purpose of searching for these new compounds was that they could be brought into the mesomorphic state at lower temperatures (in the room temperature range), which made them much more usable for practical applications. A major breakthrough came in 1973 with the introduction of cyanobiphenyl liquid crystals by George Gray because these were stable over a large temperature range.

The molecules in the liquid crystals usually have a rod-like structure. Liquid crystals can be in three phases: the nematic, cholesteric, and smectic phases (see Figure 2).

In the nematic phase the rod-like molecules are all aligned in the same direction, along a common axis, which is called the director. The term nematic refers to the Greek word νημα, which means wire or filament.

In the cholesteric phase there are layers of nematic phases, whereby the director rotates when moving from one layer to the next. The distance through the layers over which the director makes a full turn is called the pitch (just as with a screw). As the cholesteric phase is a variant of the nematic phase, it is also called the chiral-nematic phase.

The third or smectic phase was derived from the Greek word σμηγμα, which means soap or salve. The term was chosen because in this phase the liquid crystal has some mechanical properties that make it behave like soap. In the smectic phase the rod-like molecules are ordered not only in a certain direction (indicated by the director) but also in equidistant planes.

A particular type of phase that is used in displays is the twisted nematic phase (which closely resembles the cholesteric phase), first described by C. Maugin in 1911. This situation can be created because the molecules can be directed along a rubbed glass plate. When the direction in the bottom plate is different from that of the top plate, the director is twisted between the bottom and top glass plates. In displays this is combined with the phenomenon that an electric field can change the direction of the molecules due to their dipolar nature. Switching on and off an electric field disturbs the twisted nematic, and the display pixel changes from transparent to black and vice versa. This disturbance is called the Freedericsz transition, named after a Russian scientist Vsevolod K. Freedericksz, who found the effect in the 1930s. This effect is the basis of liquid crystal displays.

In the last decades of the twentieth century, numerous applications of liquid crystals have been developed. The first applications were the simple displays that we can still find in calculators and watches. For this application the nematic phase is used. Later, more sophisticated versions were developed for use in portable computers (laptops and notebooks. In this application the influence of an electric field on the molecules is used. In other applications the influence of temperature is used. This is done in liquid crystal thermometers where the cholerestic phase is required. As the reflected color depends on the pitch of the cholerestic phase, and the pitch changes with temperature, the reflected color is an indicator for the temperature when there is a mixture of different compounds with different pitches. By the end of the twentieth

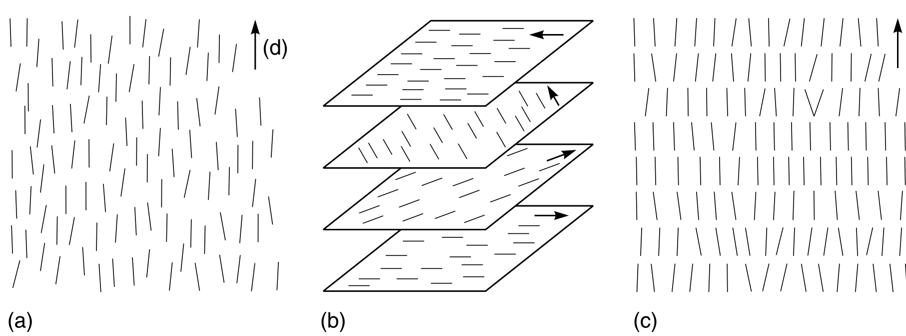


Figure 2. The phases of a liquid crystal: (a) nematic, (b) cholesteric, and (c) smectic with (d) director. [*Source: Philips Tech. Rev., 37, 5/6, p. 132, 1977.*]

century, devices for almost any temperature range were possible. Liquid crystal thermometers are used, for instance, in medical applications to measure skin temperature and in electronics to find bad connections on a circuit board by identifying places of higher temperature. Other applications include nondestructive mechanical testing of materials under stress and measuring the pressure of walking feet on the ground. Another application being explored at the turn of the twenty-first century was optical imaging. This application used the influence of an electric field, as in the display application. The liquid crystals are placed between layers of photoconductive material. At places where the photoconductor is hit by light, an electric field is produced, and this changes the state of the liquid crystal. Thus the light signal can be detected as an electric signal and then recorded.

See also **Computer Displays**

MARC J. DE VRIES

Further Reading

Bahadur, B. *Liquid Crystals:Applications and Uses*, vol. 1. World Scientific, 1990.

Collings, P.J. and Hird, M. *Introduction to Liquid Crystals: Chemistry and Physics*. Taylor & Francis, London, 1997.

Collings, P.J. *Liquid Crystals: Nature's Delicate Phase of Matter*. Adam Hilger, Bristol, 1990.

Gray, G.W. Introduction and historical development, in *Physical Properties of Liquid Crystals*. Gray, G.W., Ed. Wiley, New York, 1999.

Sluckin, Tim. The liquid crystal phases: physics and technology. *Contemp. Phys*., 41, 37–56, 2000.

Loudspeakers and Earphones

The loudspeaker is the final element of several technological systems that transmit or reproduce sounds. Once the source has been transformed into an electrical signal, the loudspeaker finally converts it back into sound waves. The earliest research on these transducers was part of the development of the telephone in the late nineteenth century. Ernst Siemens and Alexander Graham Bell received patents on dynamic or moving coil transducers in which a coil of wire moved axially within a magnetic field, driving a diaphragm to produce sound. This was the basic idea of the modern direct radiator loudspeaker, but in the era of early telephony it could only handle a small portion of the frequency range of sounds—the range that extends from the low-frequency bass notes to the high-frequency treble notes.

The introduction of Western Electric's system of electrical recording in the 1920s meant that new loudspeakers had to be devised to handle the heavier loads and extended frequency range, especially in theaters presenting the new talking pictures. Chester W. Rice and Edward Kellog of Western Electric designed a moving coil transducer with a small coil-driven cone acting as a diaphragm. This driver design was influenced by the research carried out by E.C. Wente to improve public address loudspeakers. The Rice–Kellog loudspeaker proved to have uniform response to the much wider range of sound frequencies employed in electrical recording. It was widely adopted and so completely dominated the field for loudspeakers that most other types disappeared. Rice and Kellog published the specifications of their loudspeaker in a seminal paper in 1925 and this remained the blueprint for speaker design for the next 20 years.

Loudspeakers were employed in a variety of uses in radio sets, record players, and theater sound systems. The large business organizations responsible for these products employed research laboratories to improve all elements of the system: driver, enclosure, and acoustic dampening and insulation. Bell Laboratories produced the key innovations which determined the path of development: the bass reflex principle (a vented enclosure to radiate low-frequency sounds emerging from the rear of the diaphragm) and the divided range concept, in which separate drivers dealt with different parts of the frequency range in a two-way speaker.

In the 1940s, several important designs for two-way speakers—in which the frequencies were divided between a high-frequency driver (the horn-like "tweeter") and a low-frequency driver (the large bass "woofer")—were produced for film sound reproduction in theaters. James Lansing, John Blackburn, and John Hilliard were all involved in the research sponsored by the MGM film studio, which led to the widely used two-way "Voice of the Theater" loudspeaker. Lansing and Hilliard were also involved in developing co-axial speakers, which mated small high-frequency drivers to bass woofers. These ideas were later applied to loudspeakers like the Altec Lansing Duplex 604, which established the standard for monitors in recording studios and broadcast stations and remained in production from 1943 to 1998.

During World War II, emergency research programs in electronics and acoustics produced valuable new information that was eventually applied to audio systems. The development of more powerful and responsive magnetic materials had important consequences in loudspeaker design. The new nickel-based alloys, such as alnico,

made the permanent magnet at the heart of the loudspeaker much more efficient. Wartime technological advances were applied to a broad range of products, but equally important in this progress were the young men who were trained in electronics. After the war several of these technicians founded their own innovative companies to introduce new technologies into consumer goods. Start-up companies in the audio field, such as the loudspeaker manufacturer Altec Lansing, soon challenged the technological supremacy of the large organizations.

In the 1950s, the growing interest in high-fidelity home stereo players diverted attention from movie theater sound systems, which had always been the first to incorporate new ideas in loudspeakers. The Rice–Kellog specifications determined that a very large speaker was necessary to reproduce the frequency range required by the discerning home listener. The invention of a dome-shaped tweeter by Roy Allison and the acoustic suspension enclosure by Edgar Villchur and Henry Koss removed this size constraint. Villchur used a cushion of enclosed air in the speaker to act as a stiffener, doing away with the large bass woofers needed to reproduce low-frequency sounds. The new "bookshelf" speakers, less than 60 centimeters high, set new standards for bass response. This innovation was marketed by several small companies including Acoustic Research (founded by Koss and Villchur), Advent, and Allison Acoustics.

The market for loudspeakers broadened dramatically in the 1960s: audiophiles demanded bookshelf speakers with uniform frequency response and low distortion; guitarists needed large, heavy-duty speakers to deal with the unprecedented increase of output from their amplifiers; and teenagers wanted to hear clearly the music coming from their pocket-sized transistor radios. Every user had a different concept of high fidelity, and the merits of loudspeakers were so often debated that there could be no commonly accepted standard of performance.

Earphones were only a small part of this market as they were primarily a low-fidelity alternative to loudspeakers. They were scaled down versions of the dynamic moving coil speaker, with all the same elements of the driver enclosed in a device that channeled the sound directly into the ear. The unexpected success of the personal stereo in the 1980s directed research into improving the frequency response of the earphone and establishing true stereo sound. Injecting the sound into the ear takes away the indications, such as reverberation, which give the impression of listening in a space such as a large room or concert hall. This was accomplished by programming delays into the dispersion of sound going to each ear so that the listener gets an impression of the movement of sound.

The introduction of digital sound reproducing systems placed an even greater burden on loudspeakers and earphones because the background noise and distortion was reduced and frequency response increased: listeners could hear more of the source sound and became more discriminating in their evaluation of loudspeakers. Subsequently more innovations were produced in cone and enclosure design, and in electronic manipulation and filtering of the signals in the crossover network.

The introduction of digital surround sound theater systems accelerated the pace of innovation in the 1990s, and the fruits of this research were immediately applied to home systems that employed multiple speakers and complex crossover networks that fed selected frequencies into the appropriate drivers. Innovations such as the subwoofer, which was powered by its own amplifier to give powerful bass response, brought even more fidelity and volume to the home listener.

See also **Audio Recording, Electronic Methods; Audio Recording, Stereo and Surround Sound; Audio Systems, Personal Stereo**

ANDRE MILLARD

Further Reading

Colloms, M. *High Performance Loudspeakers*. Wiley, New York, 1997.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, Cambridge, 1995.

Milton, P. Headphones: history and measurement. *Audio*, 26, 90–94, 1978.

Read, O. and Welch, W.L. *From Tin Foil to Stereo: Evolution of the Phonograph*. SAMS, Indianapolis, 1976.

Rice, C.W. and Kellog, E.W. Notes on the development of a new type of hornless loudspeaker. *J. Aud. Eng. Soc.*, 30 512–521, 1982. A reprint of the original 1925 paper.

Useful Websites

A history of the loudspeaker: http://history/acusd.edu/gen/recording/loudspeaker.html

ENCYCLOPEDIA OF 20TH-CENTURY TECHNOLOGY

TECHNOLOGY

Volume 2

M–Z

INDEX

Colin A. Hempstead, Editor

William E. Worthington, Jr., Associate Editor

ROUTLEDGE

NEW YORK AND LONDON

Also available as a printed book

see title verso for ISBN details

ENCYCLOPEDIA OF 20TH-CENTURY TECHNOLOGY

ENCYCLOPEDIA OF 20TH-CENTURY TECHNOLOGY

TECHNOLOGY

Volume 2

M–Z

INDEX

Colin A. Hempstead, Editor

William E. Worthington, Jr., Associate Editor

ROUTLEDGE

NEW YORK AND LONDON

Published in 2004 by
Routledge
29 West 35th Street
New York, NY 10001
www.routledge-ny.com

Published in Great Britain by
Routledge
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN, UK
www.routledge.co.uk

Copyright © 2005 by Routledge

Routledge is an imprint of the Taylor & Francis Group.

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's
collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage and retrieval system, without permission in writing from the publisher.

Library of Congress Cataloging-in-Publication Data

Encyclopedia of 20th-century technology / Colin A. Hempstead, editor; William E. Worthington, associate editor.
p. cm.
Includes bibliographical references and index.
ISBN 1-57958-386-5 (set : alk. paper)—ISBN 1-57958-463-2 (vol. 1 : alk. paper)—
ISBN 1-57958-464-0 (vol. 2 alk. paper)
1. Technology—Encyclopedias. I. Hempstead, Colin. II. Worthington, William E., 1948–
T9.E462 2005
603—dc22

ISBN 0-203-99780-8 Master e-book ISBN

Advisers

Dr. Jon Agar, Department of History and Philosophy of Science, University of Cambridge

Professor Janet Bainbridge, Chief Executive, EPICC (European Process Industries Competitiveness Centre), Teesside

Dr. Hans Joachim Braun, Universität der Bundeswehr Hamburg

Dr. Robert Bud, Principal Curator of Science, Science Museum, London

Dr. Michael Duffy, formerly Department of Engineering, University of Sunderland

Dr. Slava Gerovitch, Dibner Institute for the History of Science and Technology, MIT

Dr. Ernst Homburg, Department of History, University of Maastricht

Dr. Sally Horrocks, Department of Economic and Social History, University of Leicester

R. Douglas Hurt, Professor and Director, Graduate Program in Agricultural History and Rural Studies, Department of History, Iowa State University

Dr. Peter Morris, Science Museum, London

Professor John Pickstone, Centre for the History of Science, Technology and Medicine, University of Manchester

Keith Thrower, former Technical Director at Racal Electronics, UK

Contents

List of Entries

Acknowledgments

A host of workers and authors contributed to this encyclopedia, and I wish to extend my thanks to every person without whom these volumes would be stillborn. My particular thanks are offered to Gillian Lindsey of Routledge. Gillian conceived the idea of an encyclopedia of 20th-century technology, and appointed me the editor of the work in 2000. Her energy and ideas were legion, although she glossed over the amount of work for me! However, the editorship was rewarding, offering the possibility of producing a worthwhile publication with academic colleagues from around the globe. The selection of technologies and of particular subjects suggested by Gillian and me were critiqued and extended by our advisers. Their contributions, drawn from their specialist knowledge and scholarship, were invaluable. When circumstances forced my withdrawal from the active editorship, William Worthington, then with the National Museum of American History in Washington, stepped into the hot seat. To William I give my heartfelt thanks.

Finally I acknowledge the publishers and the 20th century which presented all of us with the opportunity to examine and extol some of the content and effects of modern technology. Nevertheless, the encyclopedia is partial, and any omissions and shortcomings are mine.

Colin Hempstead

My thanks go to Gillian Lindsey for presenting me with the challenge of filling the void left by Colin's departure. However, the prospect of assuming a role in a project already well under way and natural differences in approach and style were concerns. Nonetheless, the final third of the encyclopedia was crafted in such a way that it blends seamlessly with the sections completed under Colin's careful guidance. This was due in no small part to the untiring efforts of Sally Barhydt, and to her I extend sincere thanks.

William E. Worthington, Jr.

M

Mass Spectrometry

Mass spectrometry is the separation by mass-to-charge ratios (m/z) and the measurement of m/z and abundances of mono- and polyatomic ions in the gas phase. A graphical display of ion abundance as a function of m/z is called a mass spectrum. All mass spectrometers operate under greatly reduced pressure in order to produce and sustain ions in the gas phase. The British physicist J.J. Thomson observed the first mass spectrum in 1912, which consisted of differences in the trajectories of two isotopic ions of the inert gas neon, subjected to electric and magnetic fields. Near the end of the decade, Thomson's student at Cambridge University, Francis Aston, developed the first practical mass spectrograph, which was based on the apparatus used by Thomson and employed an ion-sensitive photographic plate as the detector. During the 1920s and 1930s, other investigators including A.J. Dempster at the University of Chicago and J. Mattauch and R.Z. Herzog in Austria developed mass spectrographs, and Dempster invented the electron ionization technique that is widely used today to produce ions in the gas phase. These early instruments used magnetic and electric fields to separate ions by m/z, and the discoveries made with them had a profound impact on the chemical, physical, and life sciences. The Nobel Prize in Chemistry for 1922 was awarded to Francis Aston "...for his discovery, by means of his mass spectrograph, of isotopes, in a large number of nonradioactive elements..." The atomic weights of all the elements were determined by mass spectrometric measurements of the exact masses and relative abundances of the naturally occurring and man-made isotopes.

Alfred O.C. Nier at The University of Minnesota designed and constructed improved magnetic deflection mass spectrometers during the 1930s and 1940s, and used these instruments to discover new isotopes and to measure the isotopic compositions of many elements including the uranium isotopes. Mass spectrometers were used for analyses of isotopic compositions, measurements of isotopic enrichment, and the production of uranium-235 during the World War II project to develop the atomic bomb. The Consolidated Engineering Corporation introduced the first commercial mass spectrometer in the U.S. in 1940. This instrument used magnetic deflection to separate ions and was widely employed in the petroleum industry for qualitative and quantitative analyses of gas mixtures during and after World War II. During the early 1950s Nier and his student E.G. Johnson designed a double-focusing high-resolving power mass spectrometer consisting of tandem electrostatic and magnetic sectors and the electrical detection of ions rather than photographic plates. The Nier–Johnson design was used in many commercial spectrometers, and became a standard instrument for exact measurements of the m/z of elemental and polyatomic inorganic and organic ions.

During the late 1940s and 1950s, mass spectrometer designs began to appear that were not dependent on heavy, power-consuming electromagnets. In 1955 W.C. Wiley and I.H. McLaren of the Bendix Aviation Corporation described a time-of-flight (TOF) mass spectrometer, and the Bendix Corporation marketed instruments based on their design throughout the 1960s. In a TOF mass spectrometer, ions are accelerated in batches into an evacuated flight tube by a rapid series of

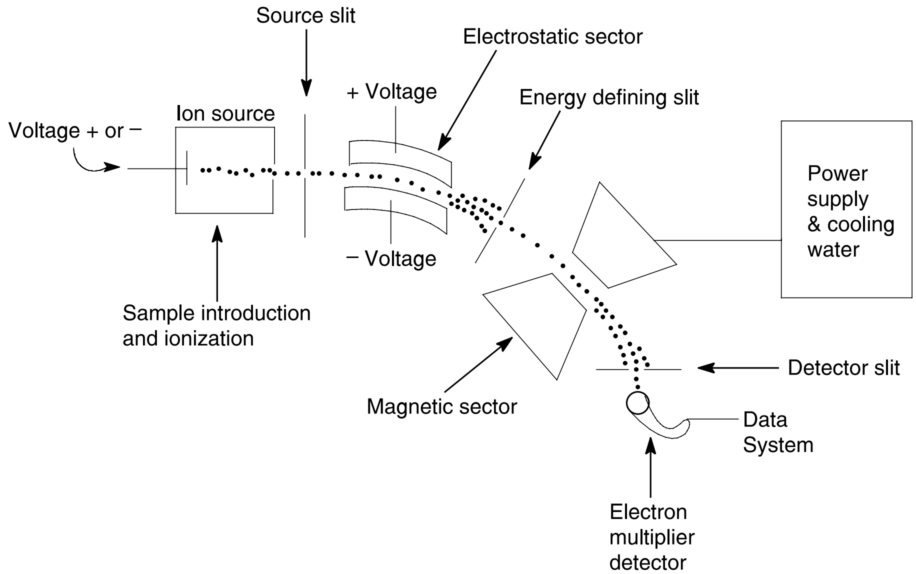


Figure 1. A mass spectrometer with electrostatic and magnetic sectors in the Nier-Johnson double focusing configuration.

high potential pulses at the acceleration electrode. The time of flight of ions from acceleration to detection is a function of m/z. However, the resolving power of vintage 1960s instruments was still too low for most applications. During the 1980s several design innovations, including the electrostatic ion mirror developed by B.A. Mamyrin and D.V. Shmikk in Russia, and very fast semiconductor electronics gave significantly improved TOF mass spectrometer performance. The TOF mass spectrometer is widely used, especially for the characterization of large molecules of biological origin and in experiments where very high-speed data acquisition is required.

Another innovation was the linear quadrupole mass spectrometer, or mass filter, described by W. Paul and H. Steinwedel of The University of Bonn in 1953. Radio frequency and direct current potentials are applied to four parallel electrodes positioned at the apices of a square and, depending on the potentials, ions of specific m/z pass through the center of the quadrupolar field to the detector while all other ions are deflected and discharged on the rods. The linear quadrupole is the most widely used type of mass spectrometer, especially when mass spectrometry is combined with gas and liquid chromatography (GC and LC) sample introduction. It is compact in size, low in weight, relatively low in cost, and has a tolerance to higher pressures—about 10^{-5} Torr—than most other types of mass spectrometers. However, it does not have high resolving power or exact m/z measurement capability. The ion trap mass spectrometer is a three-dimensional quadrupole in which ions are stored for milliseconds by a radio frequency potential applied to a ring electrode. Ions are ejected from the trap according to their m/z by increasing the amplitude of the radio frequency potential. Wolfgang Paul was awarded the Nobel Prize in Physics in 1989 for the inventions of the linear quadrupole and the quadrupole ion trap. The ion trap mass spectrometer is widely used in GC/MS and LC/MS applications.

The Fourier transform mass spectrometer (FTMS) was derived from the ion cyclotron resonance spectrometer that was introduced commercially by Varian Associates in 1966. In 1974 M.B. Comisarow and A.G. Marshall of the University of British Columbia developed the FTMS which uses a strong magnetic field, typically from a superconducting magnet, to trap ions and cause them to spiral towards the detector at circular frequencies that are related to their m/z. Unlike other mass spectrometers, ions are not physically separated, but are detected by measuring an image current induced in the walls of the trap. The m/z of the ions are determined by Fourier transforms of the image current data. The FTMS requires a very low operating pressure; that is, 10^{-8} Torr or lower, but is capable of very high resolving power and exact mass measurements.

Samples for mass spectrometry may be gases, liquids, or solids but their components must be converted into gas-phase positive or negative ions before mass analysis. Many types of sample introduction systems are used, and sample introduction and ionization may be separate or combined in a single process. Ions are injected from the sample introduction/ionization source into the

spectrometer for very rapid analysis that is typically on a microsecond or shorter time scale. Ions may be formed and injected continuously, pulsed into the spectrometer in batches, or introduced and stored in batches. Two mass spectrometers are often arranged in tandem configurations to facilitate special experiments, for example, the separation of an ion of specific *m*/*z* in the first spectrometer, collision-induced dissociation of the ion, and analysis of the decomposition products in the second mass spectrometer. All mass spectrometers are very sensitive and produce measurable signals from quantities of sample in the range of micrograms (10^{-6}) to femtograms (10^{-15}) or less. Nearly all commercial mass spectrometers produced during the last third of the century incorporate dedicated digital computers that control the operations of the mass spectrometer, acquire and store mass spectrometric data in digital form, and process or display the *m*/*z* and abundance data. Operations of most types of mass spectrometers, especially the TOF, quadrupole ion trap, and FTMS, would be impossible without very fast digital computer technology. The major application of mass spectrometry is the determination of chemical composition and structure in many fields of investigation including agriculture, beverages and foods, biological systems and processes, the environment, geology, industrial materials, petroleum exploration and processing, pharmaceutical discovery, natural products, and space exploration.

WILLIAM L. BUDDE

Further Reading

Amster, I.J. Fourier transform mass spectrometry. *J. Mass Spectrom.*, 31, 1325–1337, 1996.

Barshcik, C.M., Duckworth, D.C. and Smith, D.H., Eds. *Inorganic Mass Spectrometry: Fundamentals and Applications*. Marcel Dekker, New York, 2000.

Beynon, J.H. *Mass Spectrometry and its Applications to Organic Chemistry*. Elsevier, Amsterdam, 1960.

Budde, W.L. *Analytical Mass Spectrometry: Strategies for Environmental and Related Applications*. Oxford University Press, New York, 2001.

Cotter, R.J. *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*. American Chemical Society, Washington D.C., 1997.

March, R.E. and Todd, J.F.J., Eds. *Practical Aspects of Ion Trap Mass Spectrometry*; vols 1–3. CRC Press, Boca Raton, FL, 1995.

Montaser, A., Ed. *Inductively Coupled Plasma Mass Spectrometry*. Wiley, New York, 1998.

Nier, K.A. A history of the mass spectrometer, in *Instruments of Science: An Historical Encyclopedia*, 1998.

Watson, J.T. *Introduction to Mass Spectrometry*, 3rd edn. Lippincott–Raven Publishers, Hagerstown, MD, 1997.

Materials and Industrial Processes

The history of the development of materials during the twentieth century, as revolutionary as these developments may seem, has been part of an evolutionary past. The human race is unique among animals on this planet (except for some apes) in that we continually strive to make life easier for ourselves by inventing tools. Some essential features of all developments include the avoidance of excessive effort or extreme conditions and the achievement of improved properties as part of a war against the problems of durability, wear, rot, and corrosion. This is probably true for the development of nonferrous metallurgy since the Bronze Age. The firing of certain clays yielded pots, but many of these were porous. It was the development of glazes and particularly fluxes for glazes that led to containers that could hold fluids and avoided colonization by harmful bacteria. Fluxes could also be added to sand to reduce the melting temperature, and this led to glass manufacture. Ferrous metallurgy since the Iron Age represents a curious development. Iron is durable and malleable but not resistant to corrosion, and cast iron is brittle. However, the advantages outweighed the disadvantages and over time, techniques evolved and skills were handed down. The work of the blacksmith and specialized products such as Toledo steel have a long history.

Industrial Revolution and the Rise of Materials Science

There was a rapid acceleration of materials technology during the Industrial Revolution. Iron and steel were needed for transport (ships, trains) and in the textile industry. Iron itself became a marketable commodity; ex-catalog, made to order cast-iron shop-fronts dating from the nineteenth century can be found in Halifax, Nova Scotia. There was a need to control composition, just as there was a movement from small-batch processes to large-batch and continuous processes; the blast furnace, reverbatory furnaces, the Bessemer process and its modern derivatives. Techniques had to be developed to work the material into usable products. Hammer mills, drop forges, and rolling mills were constructed. There were improvements to lathes, drilling machines, boring and rifling machines so that they were capable of tight

tolerances over large areas of large work-pieces. But before this could be achieved in a reproducible way there was a need for standards and a uniform set of units as well as standard measures of flatness, hardness and tensile strength.

If the Industrial Revolution had major long-term effects, other developments in the Victorian era could be viewed as revolutionary in their own right, whether they were independent or spawned by the demands of increased industrialization and commercialization. The advent of the electric telegraph could be interpreted as merely the integration of the ideas of many people who were working on electricity at that time. However, it was adopted because it was needed. The provision of uniform time by which railways could be run would have been reason enough. Its value in railway signaling provided safety for the participants in the first generation of mass transport, unused to large-scale fatalities such as could occur in a train crash. The benefits for business and for the dissemination of news ensured that it was here to stay. However, the new technology demanded new materials for new applications. Subaqueous and submarine telegraphy required that the conductor be insulated and gutta percha (*cis*-polyisoprene) was found to be suitable. It is isomeric with latex rubber (*trans*-polyisoprene) and is likewise a natural product, extracted from trees that only grow in certain parts of the world. Thus it is possible to interpret an element of British Imperial strategy in the late nineteenth century as a move to protect (and control) the supply of this valuable natural resource. The demands of mass communications extended further. Improved ceramics were developed for the humble insulators that were used to support wires on overhead poles. One of the positive outcomes of the unsuccessful attempt to lay a transatlantic cable in 1857 was the unequivocal demonstration of the need for a copper conductor of the highest possible purity. Quality control was paramount and it could be argued that the ultimate success of the transatlantic venture in 1866 was largely due to the introduction of a strict system of externally monitored quality control during both the manufacture and subsequent laying of the cable.

Just as we could argue about telegraphy as a product or a mover in global developments, we could engage in a similar debate about electricity itself. High-quality copper that had been developed for telegraphy was available to provide low-loss conductors. Electric motors replaced steam and other motive sources. Energy costs for manufacturing were reduced and the demand created a new science as well as a new industry for the generation and distribution of electrical power.

Other aspects of the response to the demands that were created during the Industrial Revolution include the entire history of the chemical industry, particularly the replacement of natural dyes with synthetic dyes and the impact that this had on the textile industry. There is also the development of explosives as a means of easing the effort of mining, quarrying, tunneling and providing a strategic edge in warfare. Nitrogylcerin is easy to manufacture, but it is notoriously unstable. Blasting charges did not always undergo uniform detonation and many miners and quarry-men were killed or injured by secondary explosions. Nobel's response by integrating the explosive in an absorbent earth (kielelghur) was a timely invention that solved a problem that was seriously hampering industrial development.

Twentieth Century Technological and Consumer Needs

Materials technology in the twentieth century has been enhanced by understanding the underlying science, and tools that characterize microstructure, such as electron microscopy. As technologies required ever more stringent performance requirements of synthetic materials, novel synthesis and processing approaches have been developed. Since the 1970s, control of structural growth layer by layer (epitaxial growth in semiconductors), microstructure of composites, and targeted molecular design of polymers have made it possible to manipulate the "microstructure" to produce classes of materials with unique properties.

Fundamental developments in materials processing have helped to shape world history. There was a time when sodium and potassium nitrate, prime components in gunpowder and many explosives were in relatively short supply. During World War I, Germany did not have ready access to the natural deposits of nitrates in Chile and would have been unable to manufacture gunpowder were it not for the Haber process by which nitrogen in the air can be "cracked" and turned into ammonia (see Nitrogen Fixation). When this was converted into nitric acid, it was possible to manufacture nitroglycerin and ammonium nitrate as blasting explosives, trinitrotoluene (TNT) as a high explosive, gun-cotton as projectile explosive, as well as lead azide and mercury fulminate as detonating explosives.

The use of materials and processes to exploit latent properties to provide an easier life for more

people at reduced cost continued to operate during the twentieth century, although there has been a much greater incidence of developments for nothing more than leisure. The automobile can be seen as a device for reducing effort, but it can also be seen as a means for facilitating leisure. While the derivation of a multitude of chemicals from coal-tar (a byproduct of the coal gas industry) was a nineteenth century achievement, in the twentieth century the chemical industry has responded to the demands of a fuel-hungry mobile society by developing ever more efficient methods of maximizing the yield of fuels extracted from crude oil. This was made possible due to advances in our understanding of the mechanisms of heterogeneous catalysis (see Cracking). However, as the twenty-first century approached there was also an awareness of environmental issues. Cars are now subject to environmental restrictions. Leaded petrol is gone. There are limitations on exhaust emissions, and full recycling seems imminent.

The food industry required a protective medium for packaging and for many years cellophane (invented in 1912) fulfilled this role. Similarly, as the diffusion of electricity into domestic markets gathered pace there was a requirement for a cheap and reliable insulating plastic for plugs, plug sockets, and switches. In 1909 the response to this was bakelite, a thermosetting plastic based on phenol and formaldehyde (see Plastic, Thermosetting).

Polythene, one of the earliest plastics to come into common use, was first discovered serendipitously in 1933. Of immense importance for cable insulation during World War II, polythene is a quality insulator with a very low dielectric loss and was an ideal material for use in radar. Since then, there has been a progression from low molecular weight material obtained using extreme conditions of temperature and pressure to high molecular weight, stereo-regular polythene produced in solution using a Ziegler–Natta catalyst. The development and ubiquitous deployment of polyvinyl chloride (PVC) has almost universally replaced rubber as the insulating medium in domestic electrical installations. There have been other significant developments in the polymer industry. Polyvinyl alcohol (PVA) has replaced natural gum-arabic as a domestic adhesive and is widely used as an additive in concrete. Urea-formaldehyde and polystyrene, when manufactured as a foam, provide excellent thermal insulation used for insulating buildings. The use of epoxy resin with glass and carbon fibers to make composite materials has opened up entirely new markets. Car bodies, boat hulls and boat superstructures are regularly made using fiberglass. Carbon fiber has even better mechanical properties and is used in yacht masts, aircraft fuselages, golf club shafts, and other stressful applications. One of the more recent polymeric products that has helped to change our lives is the cyanoacrylate or superglue group of adhesives.

The twentieth century has seen the development of specialist steels at reasonable prices: chrome–vanadium steel, high-speed steel and tungsten carbide tipped tools. The cost of stainless steel has come down to a level where it is being used in applications that would have been unthinkable 50 years ago. There have also been significant developments in processing of specialist metals: titanium, zirconium, tungsten, and uranium. The addition of small amounts of thorium to tungsten was a major factor in the manufacture of low-cost, reliable electric light bulbs. The development of high-quality, low-cost aluminum and its lightweight alloys have been essential to the aviation industry.

After World War II there was a significant research effort to develop application-specific ferrite materials (ceramic ferromagnets). These were normally manufactured using the sinter process, in which a fine powder of the base material (a mixture of iron oxide with other oxides) is pressed into a mould using a wax or PVA binder. Once removed from the mould, the unit is raised to a dull-red heat and kept at this temperature for several days, during which time microdiffusion bonds the material together. Ferrites were used for rod antennas and were a key component in magnetic core memories for old mainframe computers, beginning with the Whirlwind computer in the 1950s. Ferrite films have been used as the prime data storage medium from the start of the modern computer era. They have always been used in hard disks and have continued in use as floppy disks moved from 8.5 inches to 5.25 inches to 3.5-inch high-density disks, subsequently displaced by rewriteable CDs. Layers of ferrite material deposited on thin plastic films have been used as the storage medium for tape-recorders, cassette recorders, and video tapes. The dark strip on the back of most bank, credit, and debit cards is a ferrite film that contains a magnetic record of the data relevant to the card. Specialist ferrites are now used in high-frequency transformers within switched mode power supplies, These provide more power within a smaller volume than has ever been possible before and are universally used in desktop computers as well as in the charging units for portable computers. In a curious exchange of expertise, sinter technology developed for magnetic ceramics is now used in the

manufacture of complex steel components such as gears for the car industry. There is an enormous reduction in the energy requirement when compared with cast units. The time and energy required for postprocessing is also significantly reduced.

The early part of the twentieth century witnessed an explosion in electronics. Thermionic valves (diode, triode, tetrode and pentode), or vacuum tubes, were developed to act as rectifiers, detectors, and amplifiers. They changed the nature of radio reception and made modern television a reality. Of themselves, these would have been noteworthy in world history, but they were progressively replaced after World War II by the much more reliable transistor. The twentieth century could be called the "semiconductor century." Although the selenium photoelectric cell was used in the late nineteenth century by early television scanning disk cameras (Nipkow disk) and crystalline lead sulfide (galena) was important as a crystal detector in the early days of wireless, it has been the controlled manufacture of germanium, silicon, gallium arsenide, and its relatives that has been significant. Dislocation-free crystals are required, and large diameter crystals must be grown under precisely tailored thermal gradients. Never before has any material been made in such large volumes at such high levels of purity. One of the first steps in the production of high-purity silicon involves the van Arkel process. This uses an important property in heterogeneous chemistry, namely thermal dissociation. Purified silane (SiH_4) gas is passed over a very hot tungsten wire causing it to break up into hydrogen and silicon, which deposits on the hot filament. Once this deposition has been run for some time, the rods are removed from the reactor, the tungsten is etched away and the product is ready for conversion into single crystal silicon. The situation proved to be much more difficult when efforts were made to grow single crystal gallium phosphide, the basic material for light emitting diodes (LEDs). The high temperatures needed for growth caused the more volatile phosphorous to evaporate. This presented a problem until the advent of the liquid encapsulation Czochralski (LEC) technique. Before that, the breach-end of a large naval gun was used as a pressurized reactor in the first experimental-scale production of this material.

Integration of Technologies

A key feature of recent developments has been the integration of technologies; new technology and materials frequently developed for other areas of industry were often applied elsewhere. The conversion of a semiconductor such as silicon into a useful device such as a microprocessor chip requires a wide range of technologies in the fabrication steps. Chief among these is microlithography, which provides the geometric definition (current limit for line widths is about 0.2 micrometers) of the different areas that may be the subject of a particular process step.

A slightly earlier example of the integration of technologies is the cathode ray tube in a color television receiver—the one thermionic valve that has survived when all other valves have been supplanted by discrete and integrated semiconductors. This miracle of engineering contains three electron guns, which together with their control components are connected to the outside by means of glass-metal seals. Research into the chemistry of phosphors has meant that we now have durable screens that deliver realistic colors. The ubiquitous fluorescent light is a mercury vapor discharge tube that emits ultraviolet light. The inner surface of the glass enclosure is coated with a phosphor that converts ultraviolet light to visible light. The experience gained with television screen phosphors has been incorporated into fluorescent light technology, so that the color temperature can be adjusted to the requirements of the customer.

Discharge tubes involving sodium have transformed street lighting although there is much talk of light pollution. The orange output from early sodium lights made them unpopular in shopping areas where the colors of articles on display could not be distinguished, and high-pressure mercury discharge tubes were preferred. Recent developments in high-pressure sodium tubes have provided an efficient lamp with a good color spectrum. With improvements in the design of street lighting, there has been a significant decrease in light pollution. The dissociation of gaseous compounds on hot surfaces has also been exploited to increase the light output in what are called tungsten–halogen bulbs. These are operated at much higher temperatures than normal bulbs, and they require quartz rather than the normal glass enclosures. At their normal operating temperature there is a significant "boil-off" of tungsten from the filament. This would lead to early failure except that the tungsten vapor reacts with a halogen such as iodine or bromine, which is present in the ambient gas. Since tungsten halide undergoes thermal dissociation, this mechanism ensures that the metal is quickly redeposited on the filament.

Ferrite films in magnetic storage media have increased in storage density as the methods of locating the read/write heads on the platter of hard disk have improved, but it was the integration of micro-optics for positioning that brought about the jump to multigigabyte hard drives, which by the end of the twentieth century were available at extremely low prices. In a separate application of ferrites, the integration of high-frequency electronics involving ferrites with fluorescent tube technology has resulted in low-wattage, energy-saving light bulbs. We have also seen the integration of long-chain molecules with special optical properties in array circuits where individual locations (segments) are controlled by means of transparent (tin oxide) electrical conductors in the liquid crystal display (LCD). Almost as soon as they were introduced, they caused a revolution in digital watch technology; as the twentieth century drew to a close, large-area LCDs were displacing the CRT as the display medium for television and computer monitors. As late as the mid-1970s optical fibers were a laboratory curiosity. Since then techniques have been developed to manufacture long fibers with very low losses. The integration with electronic circuits involving semiconductor lasers and high-speed photo-detectors provides the basics for a high-density communications link, the heart of the Internet era.

Just as few could have dreamt of the ubiquitous cell phone even in the late 1980s, future developments of materials and processes are anyone's guess. It is likely that tailor-made semiconductors will be prominent, but a challenge that remains is the development of improved methods of energy storage. The semiconductor manufacturers have gone a long way to reducing power requirement, but any user of a portable computer knows how limited the useful battery is. Unless there is a fundamental limit, then material scientists need to respond to the challenge of a cheap, safe and environmentally acceptable battery or battery replacement.

See also **Absorbent Materials; Adhesives; Alloys; Ceramic Materials; Chemicals; Chemicals Process Engineering; Composite Materials; Electrochemistry; Fibers, Synthetic and Semi-Synthetic; Iron and Steel Manufacture; Liquid Crystals; Nanotechnology; Optical Materials; Plastics; Semiconductors; Smart and Biomimetic Materials; Synthetic Resins; Synthetic Rubber; Thin Film Materials and Technology; Timber Engineering**

DONALD DE COGAN

Further Reading

Aftalion, F. *A History of the International Chemical Industry From the "Early Days" to 2000*, 2nd edn. Translated by Otto Benfey T. Chemical Heritage Foundation, Philadelphia, 2001.

Braun, H.-J. and Herlea, A., Eds. *Materials: Research, Development and Applications*. Brepols, Turnhout, Belgium, 2002.

Buschow, K.H.J., Cahn, R.W., Flemings, M.C., Ilschner, B., Kramer, E.J. and Mahajan, S. The science and technology of materials: an introduction, in *Encyclopedia of Materials: Science and Technology*, Buschow, K.H., Ed. Elsevier, Amsterdam, 2001.

Cahn, R.W. *The Coming of Materials Science*. Pergamon Press, Oxford, 2001.

Hummel, R.E. *Understanding Materials Science: History, Properties, Applications*. Springer Verlag, Berlin, 1998.

Mowery, D.C. and Rosenberg, N. *Paths of Innovation: Technological Change in 20th-Century America*. Cambridge University Press, Cambridge, 1998.

Medicine

Traditionally, the practice of medicine starts at the interface between the patient (or client, in late twentieth century phraseology) and the medical advisor, and that relationship should be one of mutual trust and respect. In this overview the dramatic developments in medical technology during the twentieth century are counterpointed with the changes that have occurred in the public perception of doctors.

At the end of the nineteenth century, William Osler's *Principles and Practice of Medicine* was the standard reference, and an attitude of therapeutic nihilism permeated medical thought. Cures simply could not be expected. By the end of the twentieth century, the situation was not only different, but attitudes were nearly reversed. Pharmaceutical research had resulted in powerful drugs, and complex surgery and highly technical diagnostic procedures were available. Despite the efforts of governments and individual practitioners however, the gap between the rich and the poor still existed, and not only in the poor countries. Attitudes had changed; the social structure of society had changed; patients had changed; and doctors had to change, too.

A woman born in the U.S. in 1900 had a life expectancy of 50.7 years. However, in neighboring Mexico a woman's life expectancy was only 24 years. By the end of the century female life expectancy in the U.S. had risen to 79.4 years, but in Mexico the improvement was dramatically more impressive, leading to a life expectancy of 76.5 years. Obviously something very dramatic had happened, particularly if you lived in a less affluent

country. Health care had improved in both the wealthy and less well-endowed countries, resulting in significant convergence of "outcome."

Tables listing the causes of death in a given year give an indication of the areas in which twentieth century medical technology exercised its influence. In North America in 1900 the top four causes of death were tuberculosis, pneumonia, heart disease, and enteritis (the latter being largely infantile gastroenteritis, or infections in the gastrointestinal tracts of children). By 1998 the table was headed by diseases of the heart, cancer of all forms, stroke, and chronic lung disease.

The late nineteenth century has been regarded as the golden age of microbiology, but it was nearly 40 years before there was any effective therapy for infectious diseases. Mortality from these conditions had, however, been falling for the first three decades of the twentieth century, and this can be attributed to the improvements in nutrition, public health, and a more general acceptance of the dangers of cross infection in hospital wards. In the U.K. a great debt is owed to the public works engineers of the Victorian era. Vaccination had been introduced in the second part of the nineteenth century and eventually became compulsory after the last smallpox epidemic in London in 1901. So successful was the vaccination program worldwide that in 1980 the World Health Organization (WHO) declared that smallpox was extinct.

The first antimicrobial agents had been in use as long ago as the sixteenth century when mercury was used in the treatment of syphilis. It was early in the twentieth century that German bacteriologist Paul Ehrlich introduced an organic arsenical for the treatment of that condition and noted that aniline dyes killed bacteria. The first antimicrobial agent, Prontosil (a sulfonamide) was developed in 1935 as a result of research originating in the German dye industry. This was the start of a revolution in the treatment of bacterial diseases.

In the U.K. in 1928, Alexander Fleming had observed a mold (apocryphally, blown in through an open window) that partially cleared a dish where a colony of staphylococci was growing. This "mold" was to become penicillin. Eventually the active agent was isolated and used in humans for the first time in 1941. Ernst Chain and Howard Florey started their research into what they were the first to call antibiotics, substances produced by microorganisms that kill other bacteria. By the end of the 1950s, streptomycin sulfate had been developed and tuberculosis was then curable with drugs. At the close of the twentieth century there are antibiotics to treat most bacterial infections, but progress in the development of antiviral agents has been painfully slow.

Ether, nitrous oxide, and chloroform had been introduced in Boston by Wells, and by Simpson in Edinburgh in 1846 to 1848, but it was in the twentieth century that advances in anesthesia permitted safer, longer operations. New primary general anesthetics such as halothane were safer and less toxic, but it was the introduction of muscle relaxants (developed from the curare employed by South American Indians on their arrows), mechanical ventilators, and increased understanding of pathophysiology (and the ability to both monitor and treat metabolic abnormalities) that freed the surgeons from the shackles of operating only on the younger, low-risk patients. Synthetic materials, frequently developed as a byproduct of space research at the National Aeronautic and Space Administration (NASA) led to dramatic developments in vascular surgery.

The impact of heart surgery, notably coronary artery bypass surgery, as well as the development of the artificial heart and valves are discussed elsewhere. The importance and relevance of such surgery to the prognosis of coronary heart disease cannot be overemphasized. Dramatic changes in the treatment of so-called heart attacks with drugs, angioplasty to open blocked blood vessels, and implants of stents and pacemakers have helped to increase survival rates.

Death rates from the common forms of cancer did not improve greatly in the second half of the century, and the incidence of smoking induced lung cancer had risen significantly. Cure rates for lung cancer, breast cancer, ovarian cancer, and esophageal/gastric malignancies remained distressingly low and have hardly changed although some patients had their lives extended with treatments of radiation and/or highly toxic chemotherapeutic agents. There were dramatic results in some of the less common malignancies such as lymphoma or some types of acute leukemia, notably Hodgkin's disease and some childhood leukemia. We can but hope that with further research these triumphs are the template for similar success in the more common malignancies.

After World War I an extended epidemic of influenza (Spanish flu) from 1918 to 1920 killed 25 million people, three times as many as had died in the war. In the year 2000 it was estimated that in sub-Saharan Africa alone there were 28.1 million people with HIV/AIDS infection. There were an additional 12 million people infected in Asia and the rest of the world. Without access to the retroviral drugs discovered and made available in

the last decade of the twentieth century, most of these people will die, producing a devastating demographic effect. The 1918–1920 influenza pandemic had to be faced in a state of total medical impotence, but there is, or could be, an answer to the HIV/AIDS pandemic. There are drugs to control the disease or drugs that can be given to pregnant women to prevent another generation being infected at birth. But drugs cost money, the afflicted are predominately in the less affluent world, and there are controversial issues of pharmaceutical patents and preventive methods, thus a limited impetus for action. In terms of potential global catastrophe, HIV/AIDS is the greatest challenge faced by doctors and governments. Some years earlier, tuberculosis presented a similar risk to the third world as the disease-causing organism became drug resistant. By a significant reduction in the cost of antituberculosis drugs, the threat was substantially reduced

By the end of the twentieth century organ transplantation was commonplace. The first successful renal transplant was carried out in Boston in 1954 between identical twins, so there were no immunological problems. Studies into the immunological basis of organ rejection had been carried out in the 1940s by Medawar. Once antirejection drugs (most notably azathiaprine and cyclosporin) and laboratory tissue typing had been developed, transplantation became a routine procedure. The first heart transplant was carried out in 1967 by South African surgeon Christiaan Barnard. The list of transplantable organs is now extensive, including lungs, pancreas, liver, small bowel, bone marrow, and even limbs. (Historically, corneal transplants antedate all these, but the cornea is avascular and so rejection does not occur, and tissue matching is not necessary) The problem with transplants is the supply of donor organs, which is limited for obvious reasons. As cloning became feasible there were efforts to create transgenic, cloned animals, (e.g., pigs) which could be used for human transplantation. As the century ended this work was all experimental, but it seems probable that this will be the answer to the provision of some donor tissues.

Reproductive physiology and contraception were also important areas of advance that impinged on the healthy population. Early in the century mechanical barrier methods of contraception (condoms had been used for hundreds of years) or topical spermicides were freely available. The oral contraceptives, a mixture of hormones that inhibit ovulation, were introduced in 1959 by Pincus and Sanger. They revolutionized sexual behavior and liberated women from the fear of unwanted pregnancy. Fertility was also being studied. Artificial, external insemination either by partner or anonymous donor became routine. *In vitro* fertilization using harvested donor eggs, the storage of eggs and sperm, the use of surrogate mothers, and the ability, with the use of hormone therapy, to impregnate postmenopausal women all were commonplace by the end of the century.

Psychiatric disorders are also largely amenable to treatment thanks to the very powerful psychotropic drugs developed in the last 40 years of the century. What are now considered barbaric "treatments" such as insulin comas, ice baths, and inducing malaria to cause high fever to exorcise the demons were used in psychiatric hospitals and asylums through the middle of the twentieth century and have now been relegated to the history books. There remained some level of discomfort with psychiatric disorders, which is seen in the reaction to a substratum of illness that people are reluctant to consider to be of partly psychiatric origin. This group of illnesses includes chronic fatigue syndrome, Gulf War syndrome, and several others. Attempts by physicians to suggest a nonorganic component or to fail to have a specific diagnosis and treatment plan cannot be accepted by a public accustomed to knowing the causes and finding "cures." These disabling conditions often prevent the sufferers from leading a normal life, and researchers continue to explore psychoneurological and immunological responses in an attempt to find successful treatments.

It may seem paradoxical that at a time when physicians had powerful drugs and technological treatments at their fingertips, patients should flock to consult nonmedical alternative practitioners. It has been reported that in 1990 Americans made 425 million visits to alternative practitioners (chiropractors, acupuncturists, massage therapists, etc.) but only 388 million to orthodox primary care physicians. In the latter part of the century "orthodox" physicians have, in some very basic way, been perceived to have let their patients down. There is a widespread perception that physicians have become mechanistic, do not have enough time to talk with their patients, talk down to them, and ignore what they say. Doctors could, however, learn something from alternative medicine colleagues. Some unorthodox treatments, most notably manipulative therapy, have proved as effective as high-tech neuro- or orthopedic surgery. A good orthodox physician is well aware of the valuable therapeutic effect of an unhurried

consultation with time allotted for a full explanation to the patient.

As patients expectations have mounted so has the demand for screening. There is a mistaken belief that with detailed routine screening, including a multitude of blood tests, radiological or ultrasound scans, electrocardiograms and so on, the Holy Grail of very early diagnosis will be achieved and cures guaranteed. At the end of the 1990s there were controversies related to some screening programs, notably the value of mammography screening for breast cancer and PSA (prostate specific antigen) screening for prostate cancer. The opponents of these procedures state that the very early diagnosis of breast cancer makes no difference to outcome and that PSA screening is unreliable and results in many men having prostate surgery or radiotherapy quite unnecessarily. On the other hand, screening for carcinoma of the cervix has been both successful and cost effective.

An important change in the 1990s has been the concept of evidence-based medicine; that is, treatment based on research findings of effective outcomes. Many therapeutic procedures have never been submitted to the discipline of a controlled trial. (The first controlled clinical trial was as long ago as 1754, on the treatment of scurvy with lemon juice.) Teaching medical students and other health-care practitioners in an evidence-based curriculum provides the new generation with the science and technology and the art of medicine.

The list of medical advances in the twentieth century is formidable, and many of the technologies are described in separate entries in this encyclopedia. By the close of the century medical "breakthroughs" had become a staple component of daily news. However, people born in the second half of the century who neither knew nor cared about the paucity of effective therapy before 1940 remained obstinately obsessed with the degenerative disorders such as arthritis, Alzheimer's disease, Parkinson's disease and the care of the elderly. By the end of the twentieth century, the demographics of aging populations in the western industrial countries with expectations that medical science could solve these problems meant that even these less exciting fields of research could attract grant dollars.

Memories are short, and if asked about medical major achievements of the twentieth century the questioner will be bombarded with a variety of answers mainly relating to the hot news of the 1990s: the unraveling of DNA, the Human Genome Project, the latest chemotherapeutic agent for cancer and so on. However, it is a sad fact that while the genome project is of incalculable scientific importance and interest, only a tiny number of patients with rare immune system disorders have, so far, been treated with gene therapy, with mixed results. It may be that in the future gene therapy will be helpful in the management of the wide spectrum of disorders that the genome project has already shown to have a genetic basis. In terms of direct interventional treatment any attempt to draw up a table of the most important medical advances in the last 100 years would, inevitably, be controversial and highly subjective. The development of antibiotics, insulin, and oral contraceptives would be very close to the top of most people's drug list. The science of DNA, the genome project, and the unraveling of the immune system hold much promise for the future. Coronary bypass surgery, hip replacement, and renal transplantation have benefited more people than some of the more exotic surgical introductions. Treatment of pneumonia and osteoporosis may have extended more lives than cancer chemotherapy. Consideration to both quantity and quality of life as a consequence of medical care is the norm.

See also **Cancer, Chemotherapies; Dialysis; Immunological Technology; Implants, Joints and Stents; Health; Hormone Therapy; Neurology; Organ Transplantation; Psychiatry, Pharmaceutical Treatment**

JOHN J. HAMBLIN

Further Reading

Balint, M. *The Doctor, His Patient and the Illness*. Pitman, London, 1957. Seminal work on doctor/patient interface.

Black, D. *Inequalities in Health Care*. Her Majesty's Stationary Office, London, 1980. Commissioned by one government, ignored by the next.

Dutton, D. *Worse than the Disease: Pitfalls of Medical Progress*. Cambridge University Press, Cambridge, 1988.

Ernst, E., Ed. *Desktop Guide to Complementary Medicine and Alternative Medicine*. Mosby, London, 2001.

Osler, W. *Principals and Practice of Medicine*. 1889. Classic textbook. How things were at beginning of the century.

Porter, R. *The Greatest Benefit to Mankind*. Harper Collins, London, 1997. Excellent history of medicine.

Showalter, E. *Hystories*. Picador, London, 1992. Controversial discussion of Gulf War Syndrome, ME etc.

Various. Editorials. *Br. Med. J.*, 324, 181–186, 2002. Five editorials on the HIV/AIDS epidemic.

Metals, Ferrous, *see* **Iron and Steel Manufacture**

Methods in the History of Technology

The introduction of the artifacts discussed in this encyclopedia, ranging from the automobile and the radio to the spacecraft and the computer, from wired electric lines to wireless electronic nets, and from the atomic bomb to the nuclear reactor, took place in a century more accustomed to rapid technical change than the nineteenth century—a century marked by the creation of the "machinery question" and the development of political economy to answer it. By the twentieth century, the question no longer concerned the introduction of one wave of novel machines, but had to be reformulated to analyze the historical pattern of succeeding waves of novel machines. The question could no longer be answered by political economy alone, birthed to make sense of the world of steam engines. Over the course of the twentieth century, as steam engines were already passé while new machines kept appearing, the study of the past had to be enlisted to help society understand its own relationship to technology. The twentieth century, then, witnessed the emergence and establishment of the history of technology as a distinct historical subdiscipline.

As new machines were coming at an accelerating pace, a mass faith in the equivalence between technical and social progress reposed as the twentieth century's ideological analog to medieval religious dogma. In response, the methods of the history of technology have been overdetermined by the challenge to interpret the appeal of the so-called ideology of "technological determinism." Noticeably, history of technology as such was logically impossible before the twentieth century because the modern use of the word "technology" was not established before the first decades of the twentieth century. Satisfied by how much their subdiscipline has advanced by the wise agreement to not force one definition of technology upon its members, the community of professional technology historians now agrees that the vagueness of the word "technology" is only consistent with the protean persistence of technological determinism. In the face of a definitional openness of technology, perceptions of the object of the history of technology, propositions about what its key concepts ought to be, and, correspondingly, suggestions over practices and methods have remained pluralistic and defy any easy act of subsumption under a single theoretical framework.

For convenience, we may distinguish between two historiographical periods, separated by the foundation of the Society for the History of Technology (SHOT) in 1958, with *Technology and Culture* serving as its journal of record. Melvin Kranzberg is unanimously recognized as the single most important founder of SHOT. By focusing on technological change as the outcome of "invention," individual contributions before 1958 tended to choose methods that privileged the study of individual agency over social structure. Yet, debates investigating the proper relationship between the two concerning technical change were also present in schools of thought like the "sociology of invention," known by the works of S. Colum Gilfillan, W. Fielding Ogburn, and Abbott Payson Usher. Culture was certainly the starting point for Usher, an economic historian (*A History of Mechanical Inventions*, 1929), the literary and social critic Lewis Mumford (*Technics and Civilization*, 1934), and the art historian Sigfried Giedion (*Mechanization Takes Command*, 1948); the three individuals commonly credited for being among the most distinguished contributors to history of technology's pre-SHOT period. Moreover, the interest on invention survives to the present, quite clearly as an interest in the study of the transformation of inventiveness as a socially situated manifestation of human creativity during the transition from the individual inventor's workbench to the expansive settings of industrial research.

We now know the importance of history of technology was acknowledged by several other early historiographical currents emanating from outside the U.S. Known for its sensitivity to the history of material civilization, the French *Annales* school invited historians to favorably regard the history of technology. As early as 1935, Lucien Febvre called for a tripartite methodological synthesis of a competent understanding of the technology under consideration, of a proper placement of this technology to a series, and of the appropriate move from these series to total history. Subsequent interpretations of Febvre's manifesto frequently placed the accent on the first or the third of its ingredients, thereby privileging technical and social history respectively. Technical history has been a strength of the long-lived British *Transaction of the Newcomen Society*, founded in 1920 by the Newcomen Society for the Study of the History of Engineering and Technology. The publication of an analogously focused German history of technology journal as early as in 1909 by the Association of German Engineers (now called *Technikgeschichte*) suggests that the interest in enriching engineering studies was a second strong motivation behind the foundation of the history of technology. Evidently,

a concern with proper engineering education loomed large even in SHOT's narrow constitution as a community of professional historians. Bringing along a rather spontaneous interest in the history of technology as a depository of ideas for solving contemporary technical problems and as a testimony to the scientific nature of engineering knowledge, engineers have tended to support the methodological emphasis on the history of the "heroic" moment of technological change, namely, invention. History of technology is still marked by occasional instances of unbridled antagonisms between scholars with engineering education and with those with a humanities background.

Finally, the development of the history of technology could be stimulated (or blocked) by an overarching ideological orientation. In trying to make sense of the limited development of the history of technology in the socialist societies of the twentieth century, scholars argue that underlying the crude evolutionary interpretation of Marxism emphasizing the primacy of "forces of production" was usually a hardened variant of technological determinism that left little room for historical interpretation of the society–technology relationship. Technology historians sought to advance the history of technology against the Cold War difficulties in scholarly communication through the 1968 formation of the International Committee for the History of Technology (ICOHTEC), a Scientific Section of the International Union of the History and Philosophy of Science (IUHPS), which is part of UNESCO. The agreed methodological choice of those participating in ICOHTEC was to table issues most likely to reproduce political divisions. Hence, the result was a heightened attention to subjects that appeared more technical, and, as such, less directly subjected to nationalistic interpretations.

In the 1990s, ICOHTEC published its own journal, *ICON*. In addition to the journals already mentioned, an historian wishing to publish his or her research on the history of technology can now consider two more international journals, *History and Technology* and *History of Technology*. The history of technology has by now matured enough to support special interest groups and their publications. For example, an historian interested in the history of computing technology may consider the *IEEE Annals for the History of the History of Computing*, or, for example, assuming a special interest in the history of computing with the slide rule, the American *Journal of the Oughtred Society* or the British *Slide Rule Gazette* will prove satisfactory. Another sign of the maturity of the history of technology is its sustained support by specialized museums and, with increasing frequency, even the general museums will offer particular thematic displays of interest to those in the history of technology. Also, the availability of textbooks on the history of technology, with several focused on twentieth century technology, either from a national or an international perspective, are finding their way into print. Thomas Hughes' 1989 *American Genesis: A Century of Invention and Technological Enthusiasm, 1870–1970* is already considered a classic in the field.

From the perspective of the methodologies tried, the post-1958 period can be conveniently split into two subperiods, one that ended with the mid-1980s acknowledgment of the victory of variants of "contextualism" over the couple "internalism"–"externalism," and a second, lasting since then, which started with the cautious experimentation with versions of "social constructivism." In his analysis of the methodological profile of *Technology and Culture* between 1959 and 1980, John Staudenmaier, the third (and present) editor of SHOT's journal after Melvin Kranzberg and Robert Post, found that "methodological style" already favored was "contextual" (50 percent), followed by almost identical shares of the "internalist" (17 percent) and the "externalist" (14 percent) approach, and by "nonhistorical analysis" and metastylistic "historiographical reflection" representing the remaining percentage (12 percent and 7 percent respectively). Internalism and externalism, represented the emphasis on agency and structure respectively, or, as Staudenmaier put it, technological "design and ambience." Contextualism stood for a synthetic consideration of both. If contextualism prefigured in the works of Usher, Mumford, and Giedion, internalism was dominant in the multivolume histories of technology published in Britain, France, and the Soviet Union. *The History of Technology*, edited between 1954 and 1958 by Charles Singer, E. J. Holmyard, and A. Rupert Hall, just before the emergence of SHOT, was the most influential of all.

All available historiographical balance sheets point toward an extreme variation of contextualism. Having set as its methodological standard to retrieve historical correspondence between technical relationships and social relationships, contextualism frequently stopped short after starting from either end. As the successful opening up the black box of technology proved difficult, the contribution of contextualism, according to scholars sympathetic to such thought, consisted in setting a standard that allowed the community of professional historians to evaluate what exactly

was achieved when the black box remained closed or was unpacked only partially. The acknowledged difficulty to get close to that standard was proposed as the reason for the distinct and peculiar existence of the history of technology. In his 1996 SHOT presidential address, Alex Roland stated regarding the difficulty of penetrating deep into the black box, technology historians were unique in that they were at least trying to get inside.

This difficulty generated second guesses about the value of the history of technology as a distinct subdiscipline. First came the challenge from historians who argued that the isolation of the historical study of technology from the rest of history perpetuated the risk of the spontaneous reproduction of technological determinism within the history of technology, constantly blocking the history of technology as such from providing broader historical interpretations. Far more important was the opposite challenge, that of seeking to avoid the risk of reproducing technological determinism by abandoning the attempt at broader historical interpretation altogether. As it should be expected, this challenge took the form of the invasion of the history of technology by sociology, now known as "social constructivism." Following things to their logical conclusion, social constructivists came to question the necessity of assuming a border between the inside and the outside of the black box. In the Social Construction of Technology framework (SCOT), Trevor Pinch and Wiebe Bijker's variant of social constructivism, such a border is missing as long as an artifact is subject to "interpretative flexibility" by various "relevant social groups," before "stabilization" and "closure" is achieved. In Michel Callon, Bruno Latour, and John Law's actor–network theory (ANT) variant of social constructivism, the very demarcation between nature and society is turned into a question. To SCOT's attention toward technological success and failure, ANT adds a symmetrical treatment of human actors and natural phenomena as "actants," thereby making the historian's interest in social causality altogether irrelevant.

While contextualism starts with methods to understand causes, social constructivism sees the method as an end-in-itself. While both agree on a symmetrical treatment of success and failure from a synchronic perspective, they differ in their interest for patterns. For example, the contextualist historian of the bicycle is interested in the connection between the history of the bicycle and that of the Fordist automobile whereas the social constructivist sociologist of the bicycle is interested only in the high–low wheel bicycle connection within the history of the bicycle (contrast, for example, the history of the bicycle in David Hounshell's 1984 *From the American System to Mass Production, 1800–1932* and in Wiebe Bijker's 1995 *Of Bicycles, Bakellites, and Bulbs: Toward a Theory of Sociotechnical Change*). Social constructivism has then been blamed for a constitutional inability to consider the long run. Its focus on the history of the short run has been charged with restricting the study of the transformation of technology in use. Accordingly, social constructivism has been accused by David Edgerton for conflating the history of technology with the history of invention, thereby reproducing in effect the pre-SHOT methodological emphases of internalism that cuts the technical from the social to re-invite technological determinism. In his 2003 update of the historiography of technology, Staudenmaier interprets the outcome of the contextualism–constructivism encounter positively by arguing that it has shown that critiques of technological determinism are now more forced to:

> "... take a laborious route, seeking historical contingency deep in the gears and circuits of technological design itself." *[Staudenmaier, 1992, p.170.]*

Social constructivist methods of studying technological change may be comfortably placed under the inter–trans–cross disciplinary methodologies of science and technology studies, which grew on philosophical critiques of positivism to differentiate between historical, sociological, and anthropological studies of science and technology. By contrast, authors of the contextualist school are concerned with pursuing methodologies that will facilitate the firm recognition of the discipline by the historical profession in large. Accordingly, contextualist studies of historical change tend to merge with cultural and intellectual histories of technology.

Having endorsed pluralism as their methodological principle, technology historians have experimented with sharing and borrowing from other historical subdisciplines. Their methodological focus on opening technical black boxes is now shared by business history, influenced now by schools competing with neoclassical economics, known as "evolutionary" or "institutional" economics. In pointing to habituation rather than rationality as a source of economic action, these schools of thought favor the substitution of historical study of technological "trajectories" and firm "routines" for the assumption that technological change is the automatic result of profit-maximizing

decisions on the basis of perfect market information. At the other end of the spectrum, technology historians have also developed methodological ties with labor historians, especially towards a shared focus on the puzzling issue of the relationship between mechanization, employment and unemployment, and the skilling (deskilling or reskilling) processes. The methodological ideal of this mutually beneficial interaction has been a history of workers not stripped from machines and of machines not stripped from workers.

Problematic for as long as technology was assumed to be the mere application of science, the relationship between the history of technology and the history of science has dramatically improved after the most recent generation of science historians added the study of scientific practice to scientific theory, and after the corresponding generation of technology historians started to retrieve the importance of the whole of technical knowledge instead of exhausting themselves at establishing the epistemic status of engineering knowledge. Arching toward the other end of the spectrum, where the study for the knowledge of artifacts is replaced by the study of artifacts for knowledge, history of technology has enriched its methodological apparatus by a new relationship to material culture and similarly included fields (e.g., industrial archaeology). Technology historians have been helped by their encounter with material culture to detect a contradiction in that they themselves have been methodologically studying the history of artifacts by relying on documents rather than artifacts. Against the programmatic consensus on opening the black box, the majority of technology historians continue to privilege texts over experiential sources, thereby hesitating to agree that documents are only one of the many classes of artifacts.

Joseph Corn's own reading of the *Technology and Culture* articles that Staudenmaier read for his review of the methodological transition is suggestive; preferring to write about ideas or institutions related to technical change, slightly more than half of the authors did not write about objects at all; of those who focused on the history of technical artifacts slightly more than 70 percent relied exclusively on primary and secondary textual sources, only occasionally supplemented by oral interviews; and, more importantly, only 15 percent of those considered employing some reference to material evidence. Authors focused on objects were more likely to disappear as one moves from ancient, medieval, and early modern history to the twentieth century. Scholars who focused on objects usually worked in museums and disproportionate numbers of them were trained in archaeology. As far as the methods of those who attempted to learn from things goes, Corn identified and classified five of them: ordinary looking, technical analysis, simulation, testing through use, and archaeological science.

Finally, technological determinism in the recent decades was challenged by the findings of the study of previously invisible experiences with technology: women, children, handicapped, non-whites, and non-Westerners. The list also includes the study of the relationship between technological change and ecological destruction, and the study of technological change from the perspective of use in war by a state (as opposed to its invention within an enterprising firm). Critical here is the general methodological invitation to consider technological change in consumption rather than in production, technology as changed in use rather than through invention. The issue is not one of neglecting the study of technical change at the laboratory and the factory in favor of studying the reconfiguration of technology at the world fair and the house, but on integrating the study of the two. Novel as this methodology may seem, it was suggested by Lewis Mumford, who, as Carroll Pursell has reminded us, argued against the exclusive identification of technology with tools and machines, which, in Mumford's count, had left out hearths, pits, houses, pots, sacks, clothes, traps, bins, byres, baskets, bags, ditches, reservoirs, canals, and cities. To which Pursell adds cupboards, packing cases, ship containers, trailers, and suitcases before noting that these "static containers" (Mumford) or "containers" (Pursell) are associated with domestic and agricultural work, historically the domain of women.

Pursell's list of technologies rendered invisible includes the seismic engineering that is designed into the built environment in California, the highly mechanized kitchen adjacent to the swank hotel dining room, and tableware, chairs and everything that is usually not seen as technology. A variant of the methodology aiming at seeing the invisible has been tried by looking at actors mediating between production and consumption of technologies that we would all recognize as such. Included in these actors are technological enthusiasts, from audio outlaws to computer hobbyists. The orientation to the mismatch between what a technology was thought to be and what it turned out to be in use has produced great insights on some of the technologies that defined the twentieth century—the gap between the imagined and the real uses of nuclear

energy and the rest of the crucial twentieth century technologies included in Joseph Corn's 1986 collection in *Imagining Tomorrow: History, Technology and the American Future* is a case in point.

ARISTOTLE TYMPAS

Further Reading

Braun, H.J. Current research in the history of technology in Europe. *Hist. Technol.*, 21, 167–188, 1999.

Corn, J.J. Object lessons/object myths? what historians of technology learn from things, in *Learning from Things: Method and Theory of Material Culture Studies*, Kingery, D., Ed. Smithsonian Institution Press, Washington D.C., 1996.

Cutcliffe, S.H. and Post, R.C., Eds. *In Context: History and the History of Technology—Essays in Honor of Melvin Kranzberg*. Lehigh University Press, Bethlehem, 1989, 133–149.

Edgerton, D. From innovation to use: ten eclectic theses on the historiography of technology. *Hist. Technol.*, 16, 1999, 111–136.

Fox, R. *Methods and Themes in the History of Technology*. Hardwood Academic, Amsterdam, 1996.

Hounshell, D.A. Hughesian history of technology and Chandlerian business history: parallels, departures, and critics. *Hist. Technol.*, 12, 3, 205–224, 1995.

Laudan, R. Natural alliance or forced marriage? changing relations between the histories of science and technology. *Technol. Cult.*, 36, s17–28, 1995.

Lerman, N.E., Mohun, A.P. and Oldenziel, R. *Gender and Technology: A Reader*. John Hopkins University Press, Baltimore, MA, 2003.

Levin, M.R. What the French have to say about the history of technology. *Technol. Cult.*, 37, 1, 158–168, 1996.

Lubar, S. Representation and power. *Technol. Cult.*, 36, s54–82, 1995.

McGee, D. Making up mind: the early sociology of invention. *Technol. Cult.*, 36, 4, 773–801, 1995.

Pfaffenberger, B. Social anthropology of technology. *Ann. Rev. Anthropol.*, 21 491–516, 1992.

Pursell, C. Seeing the invisible: new perceptions in the history of technology. *ICON: J. Int. Comm. Hist. Technol.*, 1, 9–15, 1995.

Roe Smith, M., Ed. *Military Enterprise and Technological Change: Perspectives on the American Experience*. MIT Press, Cambridge, Massachusetts, 1985.

Roe Smith, M. and Marx, L., Eds. *Does Technology Drive History: The Dilemma of Technological Determinism*. MIT Press, Cambridge, MA, 1994.

Roland, A. What hath Kranzberg wrought? Or, does the history of technology matter. *Technol. Cult.*, 38, 3 697–713, 1997.

Rurup, R. Historians and modern technology: reflections on the development and current problems of the history of technology. *Technol. Cult.*, 15, 161–193, 1974.

Scranton, P. None-too-porous boundaries: labor history and the history of technology. *Technol. Cult.*, 29, 722–743, 1988.

Seely, B. SHOT, the history of technology and engineering education. *Technol. Cult.*, 4, 36, 739–772, 1995.

Sinclair, B., Ed. *New Perspectives on Technology and American Culture*. American Philosophical Society, Philadelphia, 1986.

Staudenmaier, J.M.S.J. *Technology's Storytellers: Reweaving the Human Fabric*. MIT Press, Cambridge, MA, 1985.

Staudenmaier, J.M.S.J. Rationality, agency, contingency: recent trends in the history of technology. *Rev. Am. Hist.*, 30, 168–181, 2002.

Stine J.K. and Tarr, J.A. At the intersections of histories: technology and the environment. *Technol. Cult.*, 39, 601–640, 1998.

Weber, W. History of technology in Germany after 1945: institutions, methods, fields of interest. *ICON: J. Int. Comm. Hist. Technol.*, 1, 148–171, 1995.

Weber, R.J. and Perkins, D.N. *Inventive Minds: Creativity in Technology*. Oxford University Press, New York, 1992.

Wright, J.L., Ed. *Possible Dreams: Enthusiasm for Technology in America*. Henry Ford Museum & Greenfield Village, Dearborn, Michigan, 1992.

Microscopy, Electron Scanning

The scanning electron microscope (SEM) enables the imaging of surfaces at very high resolution. The SEM was a development of the earlier transmission electron microscope (TEM), and is a close analog of the conventional optical microscope but uses high-energy electrons instead of photons. This enables it to have a very much higher resolution because, for example, the wavelength of a high-energy (c.100 kilovolt) electron is less than 1/10000 that of a photon of blue light. The specimen in a TEM has to be very thin so that the electrons can pass through and be focused by electron lenses on to a fluorescent screen to produce the magnified visible image. Surfaces can be imaged directly only with difficulty with a TEM, at a glancing angle. In a SEM, an electron beam probe interacts with the surface producing backscattering and emission of secondary electrons as the high-energy electrons strike the surface; the electron beam is scanned over each part of the surface in turn (pixels), and scattered and secondary electrons for each position in the scan are detected, usually by a cathode ray oscilloscope.

The inventors of the TEM were Max Knoll and Ernst Ruska at the Technische Hochschule, Berlin in the early 1930s, while the development of the SEM started a few years later, also by Knoll after he moved to the Telefunken Company and was working on the development of television iconoscope-type camera tubes. In the course of this work he built an electron beam scanner (1935) for studying the secondary-electron emitting properties of the iconoscope targets. This apparatus (Figure 2) had many of the features of a modern SEM—the electron beam was scanned to produce a 200-line, 50-fields per second horizontal line scanning (raster), and a cathode-ray tube, deflected

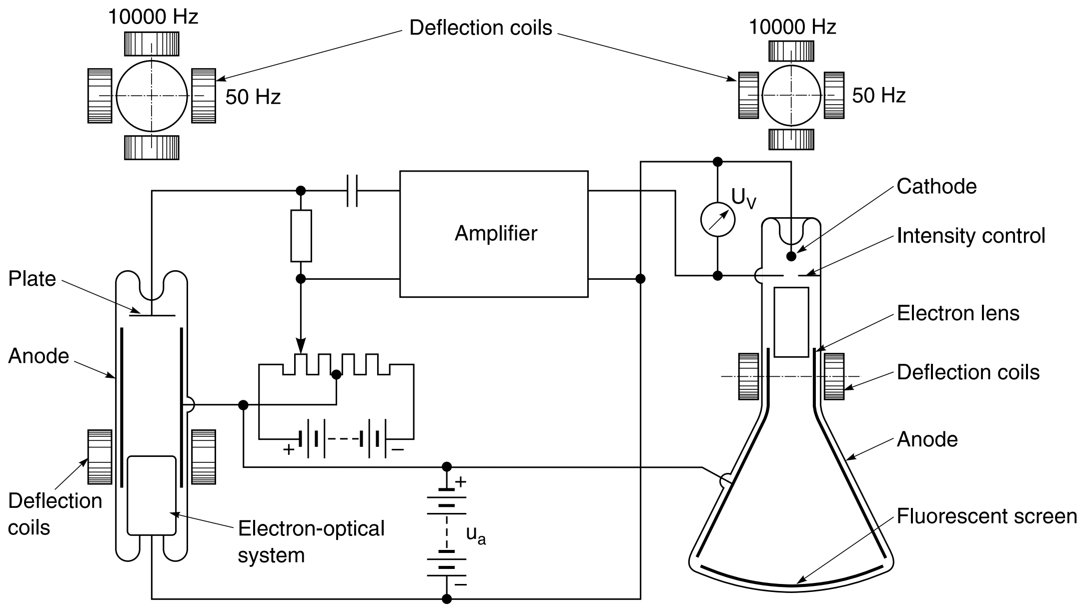


Figure 2. Diagram of Knoll's 1935 electron beam scanner.

in synchronism with scanning of the specimen and modulated in brightness by the secondary electron current from the specimen (labeled "plate" in Figure 2), displayed a visible image. This current depended on the elemental constituents of the specimen surface and, if etched, its topography; but the scanner lacked one essential feature of an SEM, a highly "demagnified" probe. In the SEM, magnification is simply the ratio of the different scanned areas at the sample and at the image. Higher magnification comes from scanning a smaller area with a sharply focused beam, made by the SEM electromagnetic lenses. Knoll was working mainly at unity magnification (or sometimes up to ten times) using a single lens to focus the beam from the electron gun on to the specimen.

The idea was extended in 1937 by Manfred von Ardenne in his private laboratory at Berlin-Lichterfelde using magnetic lenses to produce an electron probe of less than 10 nanometers (nm) diameter for a scanning transmission electron microscope (STEM), but it was unsuccessful as an SEM because he did not have a suitable detector.

A further attempt was begun at the Radio Corporation of America (RCA) in 1939 by a team led by Vladimir Zworykin (the inventor of the iconoscope). They were the first to construct a working SEM but the project was abandoned because its resolution was much inferior to that of a TEM and because of a development in surface imaging, the replica technique, which was reported in 1940 by Hans Mahl in Germany. These replicas were membranes of plastic or oxide, which bore the imprint of the etched surface being studied, and which were thin enough to be imaged in a TEM. The use of the TEM in metallurgy began from this date and for many years replicas were regarded as a satisfactory procedure for surface imaging although they were tedious to make and prone to have artifacts (false structure); even so, the consensus among most electron microscopists in the 1940s was that any further attempt to develop an SEM would be a waste of time. It required someone outside the field, Charles Oatley of the Cambridge University Engineering Department, to start the development of a microscope that could image surfaces directly and which led to the modern type of SEM.

In 1948 Oatley gave a research student, Dennis McMullan, the task of building an SEM, which resulted in a microscope with two electrostatic lenses and used an electron multiplier as the detector. Figure 3 shows a very simplified schematic diagram of an SEM from the 1950s and it is still applicable today. The components and construction of TEMs and SEMs are similar with the fundamental difference that in a TEM the lenses are used to produce a magnified image of the specimen on a viewing screen, while in an SEM they demagnify the electron source (in the gun) to produce a very small diameter scanning spot focused on the specimen. The collector receives the electrons that are reflected from the specimen

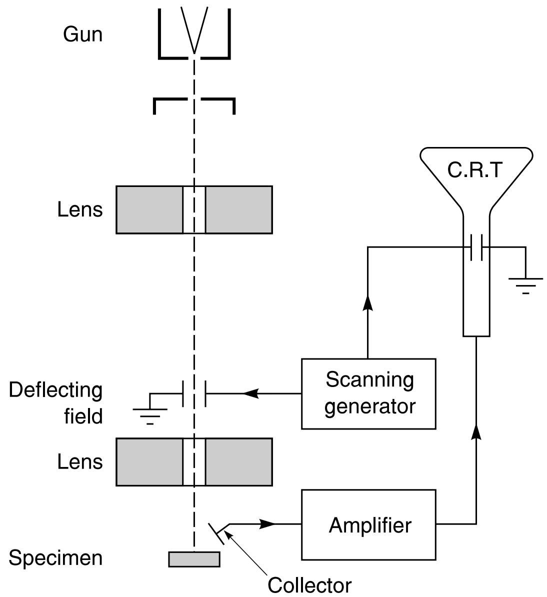


Figure 3. Simplified schematic drawing of a scanning electron microscope (1952).

Figure 4. An early visual image (of etched aluminum surface) produced with the first Cambridge SEM; 5-second scan; angle of incidence of 16 kilovolt electrons = 25 degrees; horizontal field width = 37 micrometers.

surface, and produces an electrical signal that can used to modulate the brightness of the cathode-ray tube used to display the image, as in Knoll's original electron beam scanner (Figure 2). One of the first (1952) images, of an etched aluminum surface placed at 25 degrees to the beam, shows the three-dimensional effect that is the hallmark of an SEM (Figure 4).

By the early 1960s a succession of Oatley's students had developed several SEMs that produced high-quality images, although still with substantially lower resolution than TEMs, and consequently were of little interest to the electron microscope community. However, the Cambridge Instrument Company in 1965 marketed an SEM based on Oatley's work and within a few years had sold several hundred "Stereoscans" (so-named because of the three-dimensional effect); the virtues of the SEM had quickly become apparent to electron microscopists who had actually seen one. The Stereoscan set the pattern for SEMs marketed by other manufacturers and present instruments are recognizable descendants but with many improvements due to advances in electronics, improved electron-optical components (magnetic lenses, high-intensity field-emission electron guns, etc.), and computer control.

The resolution of an SEM depends not only on the size of the focused electron-scanning spot but also, very importantly, on the scattering of electrons that penetrate the specimen and escape from a larger area of the surface, to be collected by the detector. The depth of penetration of the electrons is a function of their energy and therefore the lowest acceptable energy is used, say 1 kiloelectron volt compared with several 100 kiloelectron volts in a TEM ("acceptable" is specified because the diameter of the electron spot increases at low energies due to lens aberrations, and the current in the beam is also reduced). Advances in electron microscope technology, particularly the field-emission electron gun have enabled very low-energy electrons to be used in an SEM while still achieving good resolution. Typical spot sizes are 1 nanometer with 20 kiloelectron volt energy electrons, 2 nanometers with 1 kiloelectron volt, and 5 nanometers with 200 electron volts.

In an SEM, when the scanning beam hits the specimen not only are secondary electrons with a range of energies emitted but other radiations are as well, in particular x-rays. The energies (or wavelengths) of these depend on the constituents of the specimen surface and if a suitable detector is used, a map of the elements at the surface can be generated, which can be compared with the electron topographical image. The simplest detector is a cooled silicon diode, doped with lithium, which produces pulses proportional to the energy of the x-ray photons. Better energy resolution can be obtained with a diffracting x-ray spectrometer if

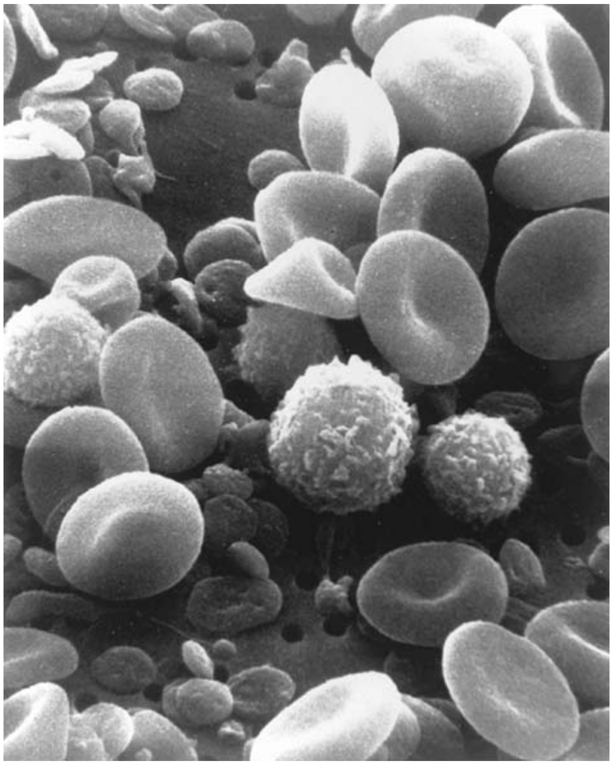


Figure 5. Scanning electron microscope image of human red blood cells and lymphocytes.
[*Source: National Cancer Institute*.]

one can be accommodated on the SEM near to the specimen. Such microanalyzers were first developed in the 1950s by Raymond Castaing in France and by Ellis Cosslett and Peter Duncumb at the Cavendish Laboratory in Cambridge: these were stand-alone instruments but later this facility was incorporated into many SEMs.

Light is also emitted by cathodoluminescent specimens and can be analyzed in an optical spectrometer; such specimens include biological substances that can be stained with suitable luminescent compounds.

Special types of SEM are used for particular classes of specimen, such as "wet" biological samples that would impair the high vacuum of a conventional SEM with water vapor. The ESEM (environmental SEM), which was first developed by Gerry Danilatos (around 1985) in Australia, operates with a specimen chamber at a relatively high pressure (around 500 pascals, still at low vacuum) so that water does not evaporate from the specimen. The column of the SEM (the region between the electron gun and the specimen) is maintained at high vacuum to enable the electron gun and lenses to operate normally.

Since the introduction of commercial SEMs in 1965, many uses have been found for them, both in advanced research and in routine analysis and imaging for science and industry. Imaging of biological cells, rock specimens, microelectronic components, and metallurgical defects can all be carried out at micrometer and nanometer scale. In fabrication of nanoscale devices and microelectro mechanical systems (MEMS), SEM imaging is an important inspection and metrological technique.

The basic components of electron microscopes continue to be improved; in particular the correction of the aberrations of the electron lenses, and no doubt the SEM will benefit from these developments.

See also **Electronics; Microscopy, Electron Transmission; Iconoscope; Microscopy, Optical**

DENNIS MCMULLAN

Further Reading

Goldstein, J., Newbury, D., Kehlin, P., Joy, D.C., Lyman, C.C., Echlin, P., Lifshin, E., Sawyer, L. and Michael, J.R., *Scanning Electron Microscopy & X-ray Microanalysis*. Kluwer, Dordrecht, 2003. The 3rd edition of a standard work on the SEM.

Oatley, C. *The Scanning Electron Microscope: Part 1, The Instrument*. Cambridge University Press, London, 1972. An early authoritative account by a pioneer.

Rasmussen, N. *Picture Control: The Electron Microscope and the Transformation of Biology in America, 1940–1960*. Stanford University Press, Stanford, CA, 1997.

Rasmussen, N. and Hawkes, P. The electron microscope, in *Instruments of Science: An Historical Encyclopedia*, Bud, R., Johnston, S. and Warner, D. Eds. Garland, London, 1997, pp. 386–389.

Reimer, L. *Scanning Electron Microscopy: Physics of Image Formation and Microanalysis*. Springer Verlag, Berlin, 1998.

Useful Websites

McMullan, D. Scanning electron microscopy 1928–1965. *Scanning*, 17, 175–185, 1995. An account of the early history of the SEM. This article is available on the web address of the Cambridge University Engineering Department: http://www.g-eng.cam.ac.uk/125/achievements/mcmullan/mcm.htm, June 2000

Microscopy, Electron Transmission

The study of electron lenses began in 1927, when Hans Busch showed that the effect of the magnetic field of a long coil on electrons can be described by the same optical law as that of a glass lens on a light beam. Soon after, Ernst Ruska in Berlin tested the familiar lens formula on a coil enclosed in an iron yoke, the first electron lens. From there,

it was a short step to combine two such lenses to form a primitive electron microscope. The first images were published by Max Knoll and Ruska in 1931, and the resolution of the light microscope was surpassed in 1933. The 1930s saw rapid developments in theoretical electron optics, notably by Otto Scherzer and Walter Glaser, and toward the end of the decade, the Siemens company began serial production. Thirty-eight instruments were built before production was interrupted by World War II.

Why should an electron microscope be better than the familiar light microscope? In the 1920s, Louis de Broglie showed that a beam of charged particles, such as electrons, has a wavelength just like a beam of light. The resolution of any microscope is fixed by the value of this wavelength, which is typically about 0.5 micrometers for light. For electrons, however, the wavelength is of the order of a few picometers (1 micrometer = 1,000,000 picometers) only, and we can thus expect to see much smaller objects with electrons.

What does an electron microscope look like? From the outside, a long metal column with various connections is all that is visible. At the head of the column is the "gun," a simple triode structure consisting of an emitter (cathode), a control electrode (known as the wehnelt), and an anode, held at 100 or more kilovolts relative to the cathode, in the case of a thermionic gun. In field-emission guns, the first electrode beyond the filament creates a high electric field at the emitter. Several lenses follow, each a rotationally symmetric electromagnet, pierced by a central canal through which the electrons pass, the yoke being terminated by a pair of high-quality iron parts known as polepieces, which create the true lens field. In one of these lenses, the objective, the specimen is introduced via an airlock. At the bottom of the column is a window though which the operator, or microscopist, can examine the fluorescent screen on which the image is formed. Outside the column are power supplies for the gun and the lenses and vacuum connections designed to reduce the pressure inside the column to about 100 micropascals (a pascal is a unit of pressure equal to 1 newton per square meter) in routine operation or 10 nanopacals if a field-emission gun is employed. The overall design is shown schematically in Figure 6.

Today, a transmission electron microscope (TEM) consists of a source, condenser lenses to direct the electron beam onto the specimen, an objective lens to provide the first stage of magnification, and several intermediate lenses and a final projector. The image is formed on a fluorescent screen for visual appraisal or on a photographic emulsion or a mosaic "charge-coupled" device (known as a CCD plate) for permanent recording. In addition, an energy analyzer is frequently incorporated in the microscope column for a type of microscopy that gives information about the chemical composition of the specimen (analytical electron microscopy); a biprism may be included for electron holography, which is a valuable technique that allows us to "see" magnetic fields directly, for example. The latest generation of microscopes may include an aberration corrector and a monochromator (see below). The power supplies to all these components are under computer control.

The "conventional" TEM forms a reasonably sharp image of the specimen (or of the diffraction pattern of the specimen in the diffraction mode), and the entire image is formed simultaneously.

In the 1960s, a new type of transmission instrument, the scanning transmission electron microscope (STEM), was developed by Albert Crewe. Here, the specimen is explored by a very fine electron probe, and the information imprinted on the probe by the specimen is retrieved by a set of collectors downstream. This usefully complements the TEM but at the cost of additional technological complexity. In particular, a very bright source is needed to keep the exposure time (the time needed to record a good image) within reasonable limits; such sources (field-emission guns) require a substantially higher vacuum than the thermionic guns that are used for routine work in the TEM.

Electron guns typically accelerate the electrons to energies between 100 and 400 kiloelectron volts, although a few very high voltage instruments have been built (3 magavolts in the Toulouse instrument, for example). At 100 kilovolts, the corresponding wavelength is around 4 picometers (1 picometer = 10^{-12} meters), but it is not possible to attain a resolving power of this order with the TEM. Electron lenses have very high aberrations and must hence be operated at a small numerical aperture, which limits the resolution to a few angströms (1 Å = 0.1 nanometers = 100 picometers). They are also very sensitive to variations in wavelength, and the energy spread of the beam emitted by the source must not exceed a few electron volts. For these reasons, the latest generation of instruments is equipped with aberration correctors and monochromators. Aberration correctors consist of sets of elements not possessing rotational symmetry. For the STEM, the device consists of a series of quadrupoles and octopoles,

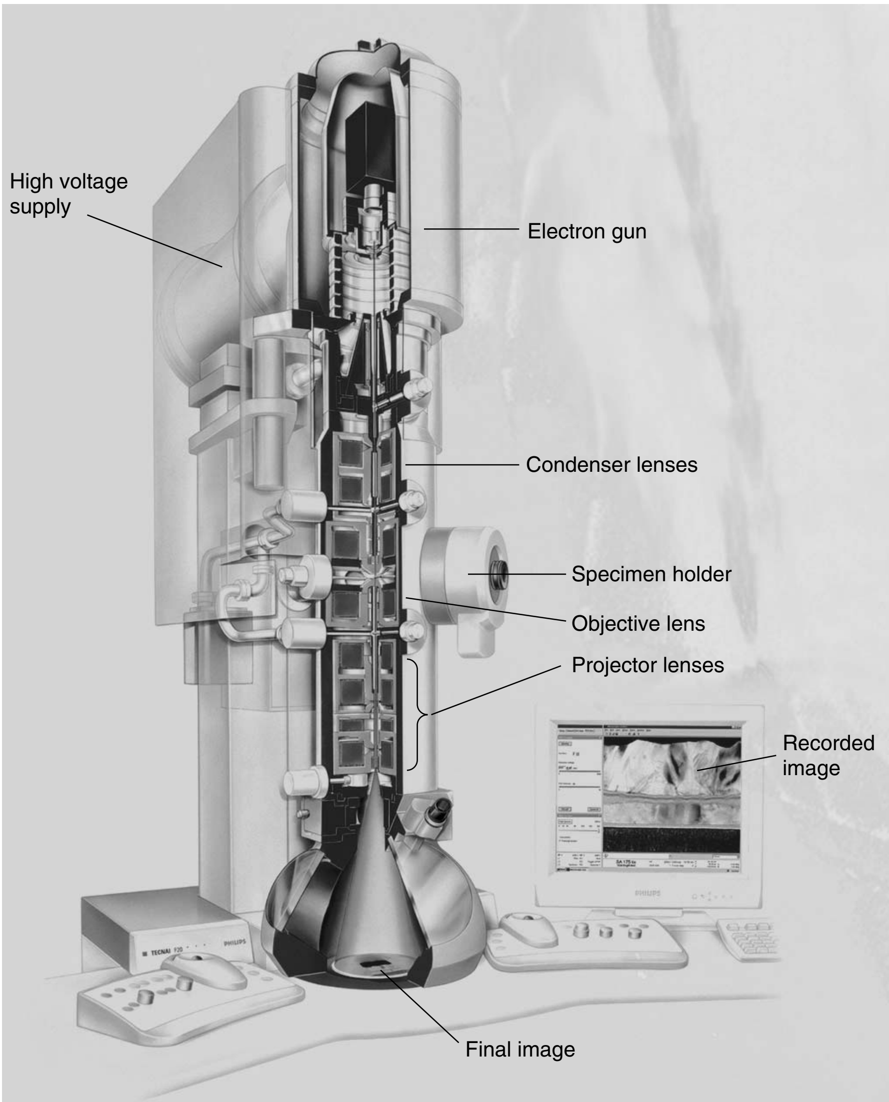


Figure 6. Transmission electron microscope.

which are capable of eliminating both the principal geometrical aberrations and the "parasitic" aberrations caused by imperfect construction or alignment. A quadrupole is the direct analog of the glass lenses used to correct the astigmatism of the human eye; there is no simple analog of the octopole, the role of which is to cancel the principal defect of electron lenses. Many STEMs are being equipped with this device today. The idea of using such optical elements was first proposed in the 1940s by Otto Scherzer, and numerous attempts were made to put it into practice before Ondrej Krivanek built a successful corrector in Cambridge in the 1990s. For the TEM, a device based on sextupoles, developed by Max Haider of CEOS on the basis of designs proposed by Harald Rose (Darmstadt) and Albert Crewe (Chicago), is preferred. As their name suggests, sextupoles consist of six magnetic poles that surround the electron beam and, once again, cancel the main defect of the microscope objective. Monochromators are devices for selecting electrons with an energy very close to the average value of the beam energy. The others are excluded so that the beam that is used for image formation is very nearly "monochromatic."

In the early years, electron microscopy developed slowly owing to the difficulty of preparing suitable specimens and the belief that the specimen would be rapidly burnt to a cinder by the electron bombardment. Gradually, however, it was realized that an electron image is formed by deflection of the electrons inside the specimen and not by absorption. In the 1950s, methods of preparing specimens thin enough to give a good image for both the life and physical sciences were found; from then on, electron microscopy has provided

invaluable information in a host of fields that could have been obtained in no other way. In the physical sciences, the relationship between the macroscopic properties of materials and their microstructure has been elucidated. The structure of magnetic materials can be seen directly. The motion of dislocations in metals and alloys can be followed in real time. Similar observations are equally valuable in the earth sciences, in polymer studies, and in such areas of applied science as catalysis and aerospace. In the life sciences, the record is just as impressive. The cell division cycle has been described in detail, and the structure of bacteria and, more importantly, of viruses has been elucidated. Thanks to the extraordinary progress of electron tomography, the three-dimensional (3-D) architecture of many biological structures has been established at resolutions on the nanometer scale. This involves the collection of hundreds of different views of the same object. After digitization, these images are processed and combined in a computer to yield 3-D images allowing the biologist to relate structure to function. There are few areas of human knowledge that have not benefited from the electron microscope.

See also **Microscopy, Electron Scanning; Microscopy, Optical**

PETER HAWKES

Further Reading

Ernst, F. and Rühle, M., Eds. *High-Resolution Imaging and Spectroscopy of Materials*. Springer Verlag, Berlin, 2003.

Haguenau, F., Hawkes, P., Hutchison, J., Satiat-Jeunemaître, B., Simon, G. and Williams, D. Key events in the history of electron microscopy. *Microsc. Microanal.*, 9, 96–138, 2003.

Hawkes, P. and Kasper, E. *Principles of Electron Optics*, 3 vols. Academic Press, London, 1996. A full, modern treatise on the optics of electron microscopes and their accessories.

Hawkes, P., Ed. *Selected Papers on Electron Optics*. SPIE Optical Engineering Press, Bellingham, 1994.

Hawkes, P. The beginnings of electron microscopy. *Adv. Electron. Electron Phys.*, 16 (s), 1985. A collection of history and reminiscence by the pioneers of the subject in many countries.

Marton, L. *Early History of the Electron Microscope*. San Francisco Press, San Francisco, 1994. An engaging personal account of the development of electron microscopy by one of the pioneers.

Mulvey, T. The growth of electron microscopy. *Advances in Imaging and Electron Physics*, 96, 1996.

Reimer, L. *Transmission Electron Microscopy: Physics of image formation and microanalysis*. Springer Verlag, Berlin, 1997.

Reimer, L. *Scanning Electron Microscopy: Physics of image formation and microanalysis*. Springer Verlag, Berlin, 1998. These two volumes provide an authoritative and well-written account of most aspects of electron microscopy.

Reimer, L., Ed. *Energy-Filtering Transmission Electron Microscopy*. Springer Verlag, Berlin, 1995.

Ruska, E. *The Early Development of Electron Lenses and Electron Microscopy*. Hirzel, Stuttgart, 1980. Original German edition, 1979. The authoritative history of the first years of the electron microscope by the builder of the first instrument.

Tonomura, A. *Electron Holography*. Springer Verlag, Berlin, 1999. The standard text on the subject by one of its greatest practitioners.

Tonomura, A. *The Quantum World Unveiled by Electron Waves*. World Scientific, Singapore, 1998. A beautifully illustrated general introduction to electron interference and holography.

Williams, D.B. and Carter, C.B. *Transmission Electron Microscopy, A Textbook for Materials Science*. Plenum Press, New York, 1996. An extremely complete and clear teaching text, covering most aspects of the subject in great detail.

Useful Websites

Royal Microscopical Society Handbooks. Many of these deal in simple language with some aspect of electron microscopy. For full details, consult info@rms.org.uk or sales@bios.co.uk

Microscopy, Optical

A microscope is an optical instrument consisting of a sequence of lenses, the purpose of which is to overcome the limit imposed by the least distance of distinct vision of the human eye, typically 25 centimeters. At this distance, the eye cannot see objects smaller than about 0.1 millimeters; 0.2 millimeters can be seen without effort. By placing a single lens, a magnifying glass, in front of the eye, the object being examined can be brought much closer and the eye then sees a distant virtual image, which is an enlarged representation of the object. The magnification M is given approximately by $25/f$, where f is the focal length of the lens (in centimeters). It was with such a primitive microscope that Anton van Leeuwenhoek made his celebrated observations of organisms in water and spermatozoa in the seventeenth century. When higher magnifications are required, a second lens is added (Figure 7); this two-lens design, with an objective lens close to the specimen and an eyepiece, is the basic form of the compound microscope, invented around 1610 by Galileo. The magnification is now given by $-(25/f)(v/u)$, where the distances v and u are defined in the figure (the minus sign indicates that the image is inverted). The design of the objective lens requires considerable care and skill since a simple glass lens suffers from defects known as aberrations, asso-

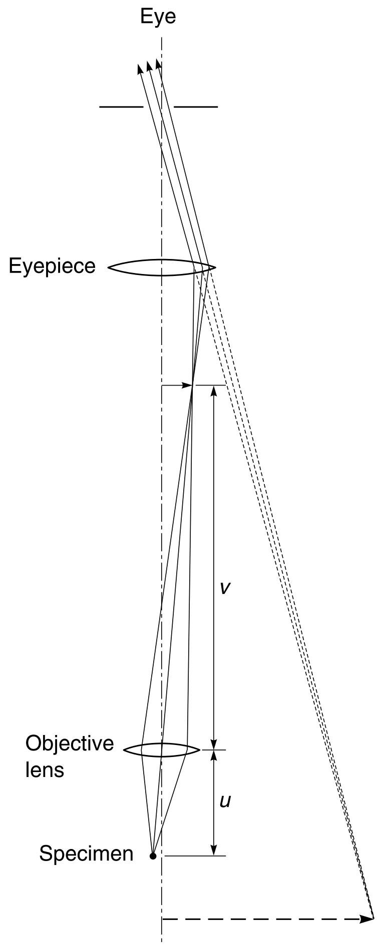


Figure 7. Basic compound microscope.

ciated with the steep inclination of some of the rays (spherical aberration, giving spatial distortion) and the spread of colors in white light (chromatic aberration). These defects can be eliminated or kept acceptably small by a suitable choice of the shapes of the lens surfaces and of the type of glass employed.

The specimen is placed on a glass slide and may be protected by a glass cover slip. The specimen stage can usually be rotated and translated. For opaque specimens, the microscope may be equipped with an epi-illuminator, a source that shines light from above the specimen; light reflected from the surface then generates the image.

It might seem from the above geometrical arguments that there is no limit to the magnification of a microscope. Although it is true that extra lenses could be added to attain very high values of the magnification, it was shown by Ernst Abbe of Zeiss in the nineteenth century that there is a limit beyond which no further fine detail can be discerned in the image. The microscope has a limit of resolution, a consequence of the wave nature of light. This limit is given by $k\,\tilde{\lambda}$ NA), in which k is a constant ($k \approx 0.6$), λ denotes the wavelength of light (about 0.6 micrometers in the middle of the visible spectrum), and NA is the "numerical aperture," a quantity that measures the light-gathering ability of the objective lens and is determined by the largest angle "seen" by the objective (the so-called angle of acceptance) and the refractive index of the medium between the specimen and the lens surface; in order to increase this refractive index, oil may introduced between the lens surface and the specimen. The numerical aperture is at best about 1.45. In practice, the resolution can reach about 0.2 micrometers with an oil-immersion objective and about 0.3 micrometers without oil. The useful magnification is hence about 1200 and 800 times, respectively; increasing the magnification will reveal no further detail.

The contrast seen in the image usually arises from absorption of the light incident on the specimen by opaque features of the latter. The amplitude of reflected light is thus reduced by varying amounts, depending on the opacity, and these variations in amplitude are perceived by the eye as the image of the specimen. However, some biological specimens are largely transparent and generate no amplitude contrast. They may nevertheless be of interest because of thickness variations or changes in the refractive index of the material of which the object is composed. Such specimens are said to alter the phase of the illuminating beam but not its amplitude and special arrangements have been devised to create "phase contrast" from such objects. In order to explain how this is achieved, we must first introduce the notion of coherence. Light sources are typically large and produce white light, which contains a broad range of visible wavelengths. Such sources are said to be incoherent because the light emitted form any point on the source is unrelated to that from any other point. At the other extreme are point sources emitting a very narrow range of wavelengths; these are both spatially (small emitting area) and temporally (narrow wavelength spread) coherent and it is such sources that are required for phase-contrast microscopy, discovered by Frits Zernike in the 1930s. Here, part of the light from the specimen passes through a phase plate, a thin layer of glass, the thickness of which is chosen in such a way that the phase of the light passing through it is altered by a suitable amount. This phase-shifted beam is recombined with the

other part of the light beam and the result is that the phase differences, invisible to the unaided eye, are converted into amplitude (or brightness) variations. For this invention, Zernike was awarded the Nobel Prize in 1953.

The basic form of the microscope may be modified and extended in many ways. Frequently, a binocular eyepiece is added, and the microscope tube may be duplicated to give stereoscopic effects. From the 1960s it became common for lenses to be coated, for example, to reduce glare. It is usual today to add a recording medium, film or a numerical device, on which individual images or sequences can be captured. In the 1970s it became possible to capture microscopical images with a television camera, and then to digitize them. Images recorded in numerical form are often transferred to a computer for subsequent analysis or processing. For the study of crystals, particularly minerals in thinly cut rock sections, the incident illumination may be polarized by means of a Nicol prism, the polarizer (a natural rhomb of calcite, cut in two parts, which are then cemented together with Canada balsam). The light emerging from the specimen is analyzed by means of a second Nicol prism (the analyzer). Polarized light microscopy (often called petrographic microscopy) gives valuable information about the boundaries between mineral grains and can be used for identification because the refractive index of many crystals is not the same in all directions (anisotropy). Polarized light selects a particular direction and, on rotating the specimen, different amounts of light will be transmitted, depending on the relative orientations of the crystal and the direction of polarization.

In a confocal microscope the optics are designed in such a way that information is gathered only from a very thin slice of the object, rejecting out of focus light from other points of the specimen. A screen with a pinhole at the other side of the lens removes all rays of light not initially aimed at the focal point. The specimen is illuminated by this spot of light and the specimen scanned point by point. There is never a complete image of the sample—at any given instant, only one point of the sample is observed and the detector builds up the image one pixel at a time. Scanning optical microscopes (SOM) had been proposed in 1928 by Edward H. Synge, and first built in 1951 by John Z. Young and F. Roberts. The performance of a SOM was improved when in 1955 Marvin Minsky invented the confocal scanning microscope. The resolution of Minsky's confocal microscope was up to twice that of a simple SOM.

The phenomenon of fluorescence has given rise to the family of confocal fluorescence microscopes. A few specimens emit light (or fluoresce) naturally but the fluorescence microscope exploits the fact that other specimens can be "stained" with fluorescent dyes, the colors of which correspond to different features of the object. Fluorescence microscopes in use today follow the basic design of Johan S. Ploem from the 1970s. Light from a laser (i.e., very high intensity) is directed onto the specimen by means of a dichroic mirror, a plate that reflects light of all wavelengths below a threshold value and transmits all wavelengths above this threshold. This incident light causes the dye in the specimen to fluoresce at a color with a longer wavelength; a small pinhole in the image plane selects the fluorescence signal, which is then recorded. Such an optical arrangement is very sensitive to the exact level in the specimen from which the fluorescence emanates. In order to form a full image of the specimen, scanning mirrors are added (Figure 8), with which the beam scans the specimen in a raster pattern (like the spot on a television screen); these mirrors also cancel the scanning effect on the return beam. The latter is intercepted by the pinhole and the signal that passes through the opening falls on the detector, which records an image of a particular layer of the specimen. By focusing on successive layers, a full three-dimensional image can be built up.

Another new member of the microscope family is the scanning near-field optical microscope (SNOM), a close relative of the atomic-force microscope, which provides information far beyond the traditional resolution limit defined earlier. The SNOM was independently developed by Dieter W. Pohl and others at IBM in Zurich and Aron Lewis at Cornell University. Here, a sharp conical optical fiber tip is placed very close to the surface of the specimen. The diameter of the tip and the distance between tip and surface are much smaller than the wavelength of light, typically 50 nanometers or less. The surface is then displaced systematically and the tip is raised or lowered to ensure that the tip–surface distance remains constant. Two signals are recorded: the intensity of the light reflected from the surface (or fluorescence excited in the sample) and the vertical movement of the tip. The first is used to form the SNOM image while the second provides topographic information about the surface. Very many variants on this basic design have been developed, notably a transmission version for transparent specimens.

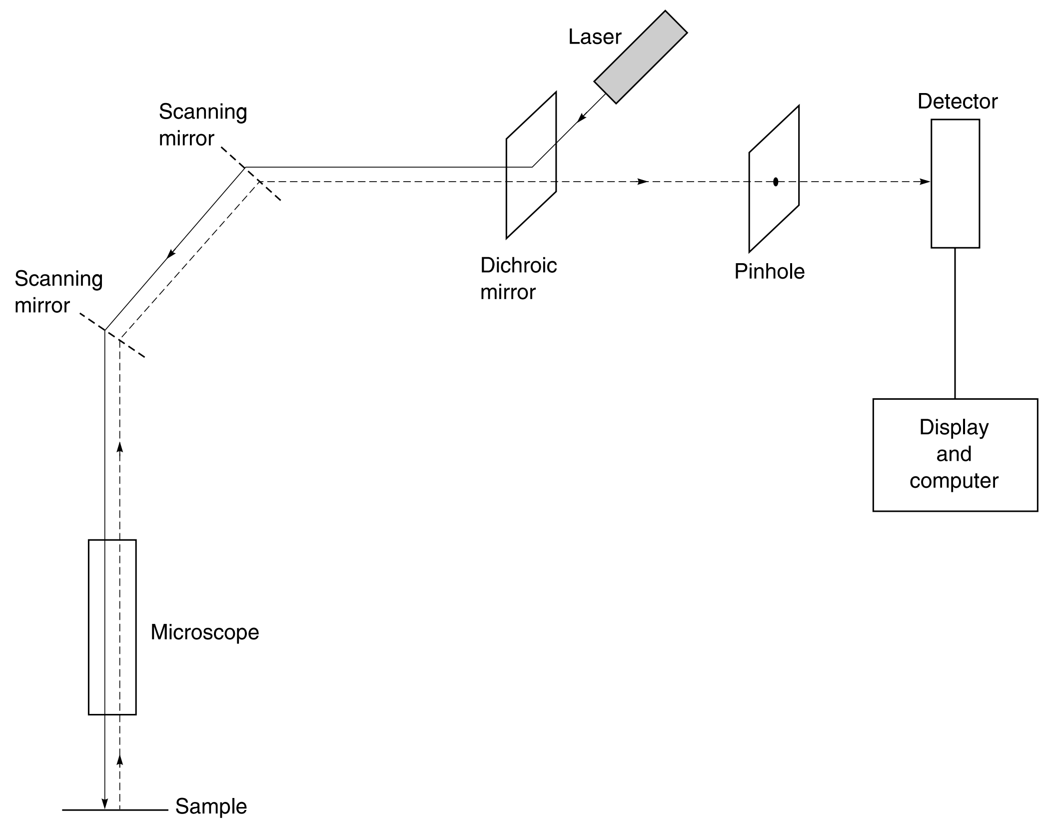


Figure 8. Confocal microscope.

See also **Microscopy, Electron Scanning; Microscopy, Electron Transmission**

PETER HAWKES

Further Reading

The series of Royal Microscopical Society handbooks published by Bios give simple introductions to most types of microscope. The most relevant volumes as well as a combined book and CD-ROM on microscopy and other titles are listed here.

Bonnell, D.A., Ed. *Scanning Probe Microscopy and Spectroscopy*. Wiley, Chichester, 2001.

Bradbury, S. and Bracegirdle, B. *Introduction to Light Microscopy*. Bios, Oxford, 1998.

Bradbury, S. and Bracegirdle, B. *Contrast Techniques in Light Microscopy*. Bios, Oxford, 1996.

Goldstein, D.J. *Understanding the Light Microscope: a Computer-Aided Introduction with CD-ROM*. Academic Press, London, 1999.

Herman, B. *Fluorescence Microscopy*. Bios, Oxford, 1997.

Murphy, D.B. *Fundamentals of Light Microscopy*. Wiley, Chichester, 2001.

Oldfield, R. *Light Microscopy: An Illustrated Guide*. Wolfe, Aylesbury, 1994.

Robinson, P.C. and Bradbury, S. *Qualitative Polarized Light Microscopy*. Bios, Oxford, 1992.

Sheppard, C.J.R. and Shotton, D.M. *Confocal Laser Scanning Microscopy*. Bios, Oxford, 1997.

Slayter, E.M. and Slayter, H.S. *Light and Electron Microscopy*. Cambridge University Press, Cambridge, 1993.

Microwave Ovens

A byproduct of the development of radar during World War II, the microwave oven is now a common part of the modern kitchen.

The microwave oven uses high-frequency radio waves generated by a magnetron tube to heat food by molecular friction. To put it another way, it causes the water molecules in the food to vibrate vigorously, and this vibration creates heat. Unlike conventional methods of cooking, where food is cooked from the outside in, the microwave oven cooks from the inside out. The high-frequency waves penetrate to the interior of the food. This results in considerably shorter preparation times than for other means of cooking. A culinary shortcoming, however, is that the microwave oven cannot ordinarily make foods, notably meats, brown and crisp.

Because of its derivation from radar, the microwave oven is an exception in the history of kitchen technology, the components of which

generally evolved from simple origins to their present level of technology; the microwave oven arrived on the scene at essentially its present state of development.

Although the microwave oven appeared after World War II, the concept of cooking with radio waves predated its development. Even before the end of the war, *Science Digest* (October 1944) predicted postwar development of a "special electronic oven" that would "employ high-frequency radio waves" for cooking. It foresaw furthermore that

> "A post-war innovation of the frozen food processors will be the completely prepared dinner. The shopper will choose between menus oered by competing companies . . . Then, one minute before dinnertime, she will place the precooked frozen meal, in its sectioned, plastic container, into a special electronic oven. This oven will employ high-frequency radio waves, which penetrate all foods equally, warming a whole chicken as fast as a portion of peas. In a few seconds a bell will ring and the whole dinner will pop up like a piece of toast—ready to serve and eat."

The prediction in *Science Digest* was apparently based on a report on the development of frozen dinners for the armed forces by the W.L. Maxson Company of New York.

Radar (radio detecting and ranging) was used for the first time by Great Britain during the Battle of Britain in 1940 and was then shared by Britain with the U.S. by then-secret agreement. The American firm that was designated to collaborate with Britain was the Raytheon Company, working in cooperation with the Radiation Laboratory of the Massachusetts Institute of Technology. Raytheon produced roughly 500,000 magnetron tubes during the war.

After the war, with demand for magnetron tubes sharply reduced, Raytheon proceeded with development of the electronic oven—what would come to be called the microwave oven—and introduced the first one to the public in 1946. Called the Radarange, it was demonstrated publicly for the first time at a press conference at New York City's Waldorf-Astoria Hotel in October 1946. By now the public was familiar with radar, and a *New York Times* headline on a story about the press conference had merely to say, "Stove Operating with Basic Radar Tube Will Cook Household Meals in Seconds."

The technology reflected in the Radarange was entirely new and, until the end of the war, highly restricted. Nearly overnight, it was not only finding its way into the public domain but being heralded as something for the most ordinary of places—the household kitchen. It was a remarkably abrupt about face. Equally remarkable was that the average housewife was thought ready to accept a device technologically so much more sophisticated than a kitchen range or any of a number of other kitchen appliances she was used to operating.

As a practical matter, however, the average housewife would not have a microwave oven in her kitchen for a number of years. The Raytheon Radarange, for the time being, was intended only for commercial use. It was roughly the size of a refrigerator, about the equivalent of five modern microwave ovens placed one atop the other. The principal reason for its size was that it required water-cooling, and hence a connection to plumbing, to keep the magnetron tube from overheating.

While it thus lacked the compactness of something for countertop use, in another respect it was easily recognizable as the progenitor of its modern counterpart: "a screened oven door, so that foods may be watched as they cook" (quoting a contemporary account). Though in full production by 1947, Radaranges were used only in restaurants, on trains, and aboard ocean liners.

By 1949 a prototype for household use was ready for demonstration, and was shown to the press at a test kitchen at Columbia University in New York City. But this second-generation Radarange was still almost the size of a refrigerator; and since the magnetron tube alone cost some $500 to produce, the household version promised to be well beyond the budget of the average home.

The household Radarange never went into production. It was not until 1955 that there was a microwave oven suitable for household use. The manufacturer was Tappan, which bought the rights to Raytheon's patent. The Model RL-1 was designed for in-the-wall installation. At 61 × 61 × 69 centimeters, it was not as compact as the average modern microwave oven, but it could be used on the countertop as well as in the wall.

The key to making it appreciably smaller than the Radarange was Tappan's development of an air-cooling system to keep the magnetron tube from overheating, a clear improvement over the Radarange's water-cooling system. The 1955 Tappan microwave nevertheless remained expensive at $1200. Only 54 were sold the first year.

Raytheon and Amana Refrigeration Inc. of Amana, Iowa, announced plans for production of microwave ovens in 1967 and other companies followed. The price was down to less than $500 by the early 1970s. It is estimated that more than

100,000 ovens were being sold annually in the early 1970s.

Improvements in methods for producing the magnetron resulted in a continual reduction in the price. As the microwave oven gained popularity in Japan in the 1970s, competition was further stimulated, driving down the price still more. By the turn of the twenty-first century, microwave ovens could be purchased for well under $100.

See also **Electronics; Radar, High Frequency and High Power; Radio-Frequency Electronics**

MERRITT IERLEY

Further Reading

Amana Refrigeration Inc. *The Amana Radarange Microwave Oven Cook Book*. Amana Refrigeration, Amana, IA, 1975.

Block, S.S. New foods to tempt your palate. *Science Digest*, October, 1944.

Ierley, M. *The Comforts of Home: The American House and the Evolution of Modern Convenience*. Clarkson Potter, New York, 1999.

New York Times. News of food: stove operating with basic radar tube will cook household meals in seconds. *New York Times*, 8 October, 1946.

Van Zante, H.J. *The Microwave Oven*. Houghton Mifflin, Boston, 1973.

Military versus Civil Technologies

The exchange of technical ideas between the military world and the civilian world can be found throughout the history of technology, from the defensive machines of Archimedes in Syracuse about 250 BC, through the first application of the telescope by Galileo in military and commercial intelligence, to the application of nuclear fission to both weaponry and power production. In the twentieth century, as the military establishments of the great powers sought to harness inventive capabilities, they turned to precedents in the commercial and academic world, seeking new ways to organize research and development. By the 1960s, the phrase "technology transfer" described the exchange of technique and device between civilian and military cultures, as well as between one nation and another, and provided a name for the phenomenon that had always characterized the development of tools, technique, process, and application.

Until late in the nineteenth century, the process of invention itself was traditionally viewed in much the same light as the process of scientific discovery. That is, writers on the topic focused on the inspired individual who, by application of intuition and intellect, solved a particular mechanical or developmental problem in his (or her) head and then worked to implement the innovation through iterative trial and error until a perfected resolution was achieved. The literature tended to present such tales as moral lessons, showing that virtues of dedication, hard work, and persistence in the face of skepticism and tradition eventually conquered obstacles, leading either to financial reward and fame or to belated recognition. More often than not, well into the twentieth century, such tales were also presented as a gloss on national virtues, with British precedents stressed in works authored by Britons, American genius dominating the works of authors in the U.S., and similar echoes of national pride found in the recorded achievements of Italians, Russians, Germans, and others.

Furthermore, the process of invention was viewed as an extension of the "great man" school of historical writing that dominated historiography of politics and statehood. Rarely did scholars look behind the biographical narrative of the scientist or technological inventive genius to try to uncover the cultural, social, or psychological roots of invention and their cross application in military and civilian spheres.

However, as technological challenges and opportunities became more complex in the late nineteenth century, the locus of achievement subtly changed from the lone genius to the team and to the accumulation of component innovations into ever more complex systems. Academics, businesses, and the military alike began to establish institutions in which programmed problems would be solved and the applications worked out in group settings. The "invention factory" established in 1876 by Thomas Edison in Menlo Park, New Jersey, although often cited in American literature as the precedent for such an approach, was part of a much larger international movement. In Britain, the Admiralty supported the construction of a model basin or towing tank by William Froude, allowing him to move from his self-financed experiments at Torquay to a staffed center at Haslar. The U.S. Navy established a smokeless powder research program at a torpedo station in Newport, Rhode Island. The German technical school at Charlottenberg established an engineering station that served as a model for an American engine laboratory built by the Navy. American land grant colleges supported agricultural experiment stations in the 1880s. Russian, Italian, British, and German institutions connected with naval research or academic institutions and sometimes funded by industrialists like Nobel, Diesel,

and the Du Ponts were all in place around the turn of the twentieth century, proliferating widely in the decade from 1900 to 1910.

Although writers and the general public may have persisted in the perception that invention was the action of the lone genius, businesses and governments alike recognized that the growing complexity of technological progress often required the application of skills from a wide variety of disciplines. They set up shops, laboratories, institutes, and project offices to foster technical creativity. Even when inventions were produced by laboratories or team efforts, however, the quest for heroes in the public mind required that Americans believe that Edison invented the light bulb while the British attributed it to Joseph Swan, that Italians believe that Guglielmo Marconi invented wireless telegraphy, and that the French, British, Americans, and Germans believe that their own scientists had single-handedly invented such component-rich devices as the photographic camera, the automobile, and the dirigible balloon. In fact, all such complex inventions, with both military and civilian applications, were the product of team invention, the exchange and cross-licensing of patents, and the international flow of purloined, imitated, and sometimes legitimately purchased technology transfer.

The world of weaponry had grown by similar combinations of individual innovation, accretion of parts, and exchange of ideas. A major improvement to naval guns came in 1851 with a design developed by John Dahlgren, an American naval officer. Reasoning that the greatest force of the burning charge was at the breech of the weapon, Dahlgren designed a gun with a thick breech around the bore and a thinner barrel further toward the muzzle. Yet to bring his device to completion required skilled craftsmen, shop workers, and metallurgists. A process developed in the 1850s by T.J. Rodman at the West Point Foundry cast the guns on a hollow core. Rodman and his crew allowed the guns to cool from the inside, greatly strengthening the inner side of the barrel.

A secure, screw-breech system for loading artillery was first patented by B. Chambers in 1849 in the U.S. With the development of smokeless powder in the 1890s and with the discovery of the ideal formula for the material by Dmitri Mendeleev, designers sought to improve the breech loading system to allow for rapid firing and reloading, and several rapid-fire designs came into use after 1895. At least five different rapid-fire breech-loading designs were developed in Germany, Britain, Russia, and the U.S. in the 1890s. Although some were named after their individual inventors, all were the product of group efforts both in machining and in testing. Only in the twentieth century, however, did most armies and navies begin to designate their weapons by "Mark" numbers, sometimes retaining an individual name along with it. Thus, the U.S. introduced the M-1 rifle in the 1930s, but also designated it the "Garand."

Some authors have continued to argue that most inventions come from the individual or at most, the small firm. In a study produced in 1969, John Jewkes of Oxford University detailed the case histories of 75 twentieth century inventions, ranging from acrylic fibers through to the zipper, and showed that a large proportion of them were conceived and developed by individuals or small firms, not by industrial or military facilities. The focus of Jewkes' work tended to be civilian technologies such as the antibiotic penicillin and the gasoline additive tetraethyl lead, rather than devices conceived for military purposes.

Behind the persistent public mythology of the lone heroic inventor, the research and development laboratories of the technology corporations, the naval establishments, the armories, and the government-funded experiment stations sought means to program or schedule the process of invention. For some it seemed unlikely that the head of an organization could "order" progress. Alexander Fleming, the British researcher who accidentally discovered penicillin in 1928, believed that "a team is the worst possible way" of conducting research.

Thomas Midgely, however, who tracked down tetraethyl lead to add to gasoline as a means to eliminate engine "knock," was always ready to describe how his discovery, although accidental, came out of a funded, directed, and tedious process in which a team employed by C.F. Kettering of Delco explored alternate chemicals over the period 1919 to 1921. Kettering recognized that the knocking of early four-stroke internal combustion engines varied with fuel and that some chemical additive might reduce or eliminate the problem. In effect, Midgely produced a discovery (and an invention, in the process of making tetraethyl lead), on order. Kettering himself was a firm believer in putting experts together and tying the efforts of academically trained specialists to the practical experience of mechanics, tinkerers, and experienced craftsmen. Setting up such a group and then giving them a problem to solve was the essence of the new invention-on-order system emerging in twentieth century laboratories and workshops.

Penicillin, although discovered by Fleming individually, was not produced as a medicine until nearly 15 years after its discovery, and this was only through the hard work of a team working in the period from 1939 to 1943. That group included Australian-born Howard Florey and German-born Ernst B. Chain who found a way to produce purified penicillin, the active ingredient in the mold that Fleming had identified. Florey and Chain worked initially at Oxford University. Chain identified the chemical structure of crystalline penicillin and identified four different types. From 1941 to 1943, Florey worked with staff of the Research Laboratory of the U.S. Department of Agriculture in Peoria, Illinois, developing methods for production. Fleming, Florey, and Chain shared the 1945 Nobel Prize for Physiology or Medicine for the complete process of discovery and isolation of the antibiotic. In this case, the interaction of the military and civilian spheres was demonstrated in the vastly increased need for a drug to combat infection and disease brought on by World War II. Florey himself first worked on field tests among wounded combat victims in Sicily and Tunisia.

The identification of talent, establishing of research direction (or definition of a research problem), and management of the team effort have all presented difficult issues for industrial, academic, and military laboratories and their research and development (R&D) activities. Since interdisciplinary work is often required among people with training in varied fields such as chemistry, mechanics, computer technology, materials sciences, and others; merely structuring a team and managing it may present difficulties. Establishing a shared vocabulary sometimes requires team members to break out of their disciplines and learn to convey specialized data and concepts in the language of generalists.

One common solution to the organizational issue has been to house the specialists in departments with a disciplinary focus and then to assign them to projects on a temporary basis, in a so-called matrix organization that exists for the length of the project or development. An extension of this principle in the late twentieth century was to draw specialists from entirely different organizations in academia, the military, and in industry to work on the same project while they collected their salaries from the home organizations. In such situations, "integrated project teams" (IPTs) were created for the duration of a developmental project. IPTs proliferated in military technology development efforts in the 1990s.

Although cases of technology transfer from civilian applications to the military and vice-versa can be identified, the creation of new technology specifically for a military application is very rarely a simple matter. Converting the civilian discovery of nuclear fission to a workable nuclear weapon occurred fairly promptly. Fission was identified in December of 1938, and the first test of a device that could be fitted into a weapon case was held in July 1945. The Manhattan Engineer District, formed in 1942, consolidated work at numerous civilian and military facilities into a single project, managed as a large enterprise, all conducted behind a screen of security. Although penetrated by Soviet agents, the American project remained unknown to the Germans and to the Japanese until the first weapon was detonated over Hiroshima in August 1945. Even so, it involved tens of thousands of construction workers and hundreds of engineers and scientists, and it was conducted at research and production facilities scattered across the U.S. and Canada.

The Manhattan Project has often been regarded as the first case of "big science" in the U.S., although prior projects to build a cyclotron at the University of California in the 1930s, to construct a 200-inch (5-meter) telescope for Mt. Palomar in Pasadena, the Soviet and German efforts to construct a nuclear weapon, and the German project to build the V-2 rocket at Peenemunde were precedents or contemporary in nature. What characterized the so-called big science projects was massive funding, the organization of hundreds if not thousands of specialists, and the pursuit of a specific technological goal, all on a massive budget.

Despite such vast projects involving many different specialists, the persistence of the "great man" mythology elevated the administrators and lead scientists of such projects to the rank of historical figures. Thus J. Robert Oppenheimer became known as the "father of the bomb," while similar roles were attributed to such science and technology administrators as Werner Heisenberg, Igor Kurchatov, and Wernher von Braun.

Even the concept that such massive projects represented big science was itself a matter of debate in later years. Were such applications of scientific method actually scientific endeavors, or were they the work of engineers? In fact, the task of building the nuclear weapon represented a case of the application, not the discovery, of scientific theory, and the design of machinery and processes to build a working weapon. Although there was no such field as "nuclear engineering" in 1938, the work

done by the teams at the Metallurgical Laboratory at the University of Chicago, at Hanford and Oak Ridge in the U.S., and at Montreal and Chalk River in Canada were all engineering tasks. Scientists working as engineers and cooperating with industrial engineers and chemical engineers from such firms as DuPont collaborated with civil engineers from the U.S. Army Corps of Engineers. Together they designed production reactors to make plutonium and separation plants to isolate fissionable uranium-235 from the more plentiful isotope uranium-238, and they built the "gadgets" themselves. Regarded as a triumph of science and the work of notable physicists (Oppenheimer, Leo Szilard, Enrico Fermi, Niels Bohr, Neddermeyer, Eugene Paul Wigner, and others), in fact the weapon and the work involved in its creation was an immense engineering task.

In the U.S. following World War II, the administrator of the Office of Scientific Research and Development (OSRD), Dr. Vannevar Bush, published a work that was highly influential in capturing the sense that "science had won the war." *Science, The Endless Frontier*, published both as a report on the work of the OSRD and then as a popular work to build political support for a continued effort to fund science, made the clear argument that science had to precede application and that research had to precede development. For a generation in the U.S., Britain, and Canada, funding for military advances was directed to "research and development" projects. In fact, most of those projects represented the application of existing technology and established science rather than efforts to fund pure, basic, or abstract science, as advocated by Bush. Even so, engineers were baffled by the emphasis in the popular press and in the minds of government administrators on "rocket science" and on the vocabulary that insisted that research preceded science.

In the late 1960s and into the 1970s in both the U.S. and Britain, a lively discussion emerged over the sources of invention that challenged the Bush paradigm of science leading technology toward innovation. The *Project Hindsight* report in 1969 by Chalmers Sherwin and Raymond Isenson captured many criticisms of the Bush paradigm. The report by Sherwin, a preliminary version of which was published in the journal *Scienc*e in 1967, stirred up a hornet's nest of responses and letters. A follow-up study by the Illinois Institute of Technology, known as the *TRACES Report* in 1969, demonstrated the ultimate scientific basis for many technological advances. However, the debate continued in both the U.S. and Britain, leading to close studies by I.C.R. Byatt and A.V. Cohen in 1969, M. Gibbons and R.D. Johnston in 1972, J. Langrish in 1972, and F.R. Jevons in 1976, all of which concluded that very few important recent inventions could be attributed to advances in science.

The issue was no sterile debate between the two professions; it had serious implications for the funding of military R&D in both Britain and the U.S. In the U.S., the compromise was typically bureaucratic. In the military appropriation budgets, the Defense Department established a continuum from Basic Research, funded in a budget category or "budget element" 6.1 in the military appropriation budgets, through Applied Research (6.2), Advanced Technology Development (6.3), Demonstration and Validation (6.4), Engineering and Manufacturing Development (6.5), Management Support (6.6), and Operational Systems Development (6.7). Although the categories evolved over the period from the 1960s through the 1990s, by 1993 the pattern persisted. However, the 6.1 category of Basic Research received a very low proportion of defense budgeting, with far greater amounts proposed (and funded) in each budget cycle for categories 6.3 through 6.5. Although science was given priority of place in the intellectual scheme, in the practical world the dollars went into the costly work involved in actually building weapons and devices rather than maintaining the scientist at his bench in the laboratory.

This budgetary scheme reflected the historical reality. When Lise Meitner and her nephew Otto Frisch conceived of the idea of nuclear fission in 1938, they did so over the telephone between Stockholm and Copenhagen, while on Christmas vacation away from any expensive facilities. Seven years later, the Manhattan Project had spent 2 billion dollars to construct an industry and a device implementing the concept. Despite the fact that pure science tends to be less expensive than the construction of weapons or weapons platforms, a continuing debate remains over the degree to which basic research budget categories within the military appropriation request should be expanded.

See also **Engineering, Cultural, Methodologicial, and Definitional Issues; Engineering, Production and Economic Growth; Globalization; Organization of Technology and Science; Research and Development in the Twentieth Century; Social and Political Determinants of Technological Change; Warfare**

RODNEY P. CARLISLE

Further Reading

Bush, V. *Science, The Endless Frontier*. Washington D.C., 1945.

Bush, V. *Modern Arms and Free Men: A Discussion of the Role of Science in Preserving Democracy*. Simon & Schuster, New York, 1949.

Byatt, I.C.R. and Cohen, A.V. *An Attempt to Quantify the Economic Benefits of Science Research*. Department of Education and Science, Science Policy Studies no. 4, London, 1969.

Gibbons, M. and Johnston, R.D. *The Interaction of Science and Technology*. Report to the Council for Scientific Policy (U.K), December 1977.

Illinois Institute of Technology. *Technology in Retrospect and Critical Events in Science* ("TRACES"). Illinois Institute of Technology, Chicago, 1968.

Jevons, F.R. The interaction of science and technology today, or, is science the mother of invention? *Technol. Cult*., 729–42, 1976.

Jewkes, J., Sawers, D. and Stillerman, R. *The Sources of Invention*. 2nd edn. W.W. Norton, New York, 1969.

Langrish, J. and Gibbons, M., Evans, W.G. and Jevons, F.R. *Wealth from Knowledge*. Wiley, New York, 1972.

Rhodes, R. *The Making of the Atomic Bomb*. Simon & Schuster, New York, 1986.

Sherwin, C.W. and Isenson, R.S. *Project Hindsight: Final Report*. U.S. Department of Defense, Office of the Director of Research and Engineering, Washington D.C., October 1969.

Missiles, Air-to-Air

Interest in air-to-air missiles (AAMs, also known as air intercept missiles or AIMs) was initially prompted by the need to defend against heavy bombers in World War II. Unguided rockets were deployed for the purpose during the war, but the firing aircraft had to get dangerously close, and even so the rockets' probability of approaching within killing range of their targets was poor. Nazi Germany developed two types of rocket-propelled missiles employing command guidance and produced some examples, but neither saw service use.

Wartime work on fire control for aerial gunnery had clarified many of the guidance issues. It rapidly became apparent that there were two principal avenues of promise:

1. Semiactive radar
2. Passive infra-red (IR).

In semiactive homing the interceptor aircraft that had fired the missile kept its radar trained on the target and the missile homing head incorporated a radar receiver that received the reflected radar energy and made its own computations of best course to intercept. The radar was normally in x-band (i.e., near 9 gigahertz frequency) for good precision. IR homing heads, operating in the 1–2 micrometer band, detected the very hot jet efflux and thus were effective only for attack from astern.

AAMs were generally low-aspect vehicles with cruciform wing and tail or canard surfaces. As thermionic valves (vacuum tubes) had to be used, electronics were avoided or drastically simplified and electromechanical systems were preferred where possible. Early AAMs were bulky, expensive, and unreliable; interceptors generally carried mixed armaments of missiles together with unguided rockets or cannon.

A major milestone was the development of the IR-guided AIM-9A Sidewinder missile by the U.S. Naval Ordnance Test Station in China Lake, California. Every effort was made to find clever and simple solutions in order to reduce complexity, cost, and size. Moreover, the Sidewinder was the first AAM to show a capability in fighter-to-fighter combat. It was limited to firing positions in a narrow cone no more than about 3 kilometers behind the target fighter, but this nevertheless represented a major advance over what was possible with cannon. It received its first combat test in fighting between Taiwanese and Chinese forces in 1958, when the American-equipped Taiwanese showed significant superiority. In a succession of variants the Sidewinder became the standard short-range missile not only for the U.S. forces but those of many American allies as well. Highly evolved versions remained in wide service at century's end.

By the 1960s, many fighters were dispensing with cannon altogether and relying entirely on AAMs. This approach came under test in the conflict between the U.S. and North Vietnam (which relied on Soviet-supplied arms) in the late 1960s. In addition to Sidewinder, the American planes carried the larger semiactive radar AIM-7 Sparrow which, in principle, had a wider engagement envelope, and of course offered the attraction of operation in clouds. In early innings, however, only about 5 percent of the Sparrows found their targets. There were a number of reasons, of which inadequacies in Sparrow performance and reliability played a significant role.

These early combat experiences led to a reversion to cannon armament for fighters, but at the same time prompted intensive efforts to develop improved missiles. It was found that IR seekers were vulnerable to locking onto the sun or its reflections from water, glass, metal, or even clouds. Vietnam experience also served to heighten awareness of the difficulty in attaining an ideal firing position in combat with an agile fighter and the consequent need for the AAM to have a much

wider engagement envelope. Two other issues that had not thus far been particularly troublesome in combat emerged as a result of tests. First was the need for the fighter to keep its radar and thus its nose pointed toward its adversary throughout the flight of its semiactive homing AAMs. In a head-on engagement, this forced it to close to dangerously close ranges to ensure a hit, even with relatively long-ranged AAMs. It also was appreciated that both radar and IR missiles were potentially vulnerable to various sorts of countermeasures and decoys.

The increasing use of solid-state and ultimately integrated digital electronics made it possible at once to cut the weight and cost of guidance and control functions and increase their reliability, resulting in substantial opportunities to improve AAMs for those who could afford the necessary technology. These and other new technologies were applied to making improved versions of Sidewinder and Sparrow, which gave a good account of themselves in the Gulf War of 1991.

A new generation of AAMs was also developed, of which the American AIM-120 AMRAAM (advanced medium range AAM) is among the best known and most widely employed. AMRAAM navigates to a preset terminal point, which may be updated during flight by coded radio command from the launching fighter. On reaching the terminal point it turns on a self-contained radar, acquires the target, and homes to an intercept autonomously. In the final eight years of the century, this weapon was used over Iraq and the former Republic of Yugoslavia, reportedly with good results.

Only American AAMs have seen wide combat service, which accounts for their prominence in this account. A number of other nations have been active in AAM development, however. Reportedly, the Soviet Union was particularly vigorous in developing so-called "high off-boresight" weapons, meaning AAMs that can be successfully employed even when the nose of the firing fighter is pointed far away from the target. Advanced versions of the Sidewinder and comparable missiles of other nations were planned to have similar capabilities.

The development of stealthy aircraft posed special challenges for AAMs: how is an AAM to home (and fuse) on a very stealthy target? And how is a stealthy aircraft to employ AAMs without compromising its stealth? The technology and techniques being applied to these and other issues remain shrouded in secrecy, ensuring that the full story of twentieth century AAMs will not be known for many years to come.

William D. O'Neil

Further reading

Dow, R.B. *Fundamentals of Advanced Missiles*. Wiley, New York, 1958. Many fundamentals remain relevant, in addition to its value in documenting earlier technologies.

Fleeman, E.L. *Tactical Missile Design*. AIAA, Washington D.C., 2001. Comprehensive engineering textbook at introductory level.

Frieden, D.R., Ed. *Principles of Naval Weapons Systems*. Naval Institute Press, Annapolis, 1985. Extensive material relating to AAMs at modest level of technical complexity.

Gunston, B. *The Illustrated Encyclopedia of the World's Rockets and Missiles*. Crescent Books, New York, 1979. Capsule histories of developments from World War II through 1970s.

Ivanov, A. Radar guidance of missiles, in *Radar Handbook*, Merrill I. Skolnik, Ed., 2nd edn. McGraw-Hill, New York, 1990.

Lennox, D., Ed. *Jane's Air-Launched Weapons*. Jane's Information Group, Coulsdon, 2000. Earlier editions provide added information on earlier developments.

Pocock, R.F. *German Guided Missiles of the Second World War*. Ian Allan, London, 1967.

Whitford, R. *Fundamentals of Fighter Design*. Airlife Publishing, Shrewsbury, 2000. Useful brief summary of AAMs in pp. 189–203.

Missiles, Air-to-Surface

Precision attack of ground targets was envisioned as a major mission of air forces from their first conception, even before the advent of practicable airplanes. Until the 1970s most air forces believed that this could be best accomplished through exact aiming of cannon, unguided rockets, or freely-falling bombs, at least for most targets. But although impressive results were sometimes achieved through these methods in tests and exercises, combat performance was generally disappointing, with average miss distances on the order of scores, hundreds, or even thousands of meters.

Even moderate accuracy required the aircraft to approach the target closely, exposing it to defensive fire. In World War I, Germany developed a family of gliders guided by commands transmitted by unreeling electrical wires and intended to deliver a torpedo while the aircraft remained several kilometers distant. The war ended before these saw service.

Late in the 1930s, Germany again took up development, now focused on precision attack of ships, bridges, tunnel entrances, and other difficult targets. Glide and rocket-propelled bombs guided by radio command were employed successfully from mid-1943. Comparable weapons were also employed by U.S. services. The need for command guidance limited the launching aircraft's standoff

range and exposed it to defensive fire. Moreover, the U.S. and the U.K. deployed jammers to interfere with the radio command links of the German missiles. Just prior to the end of the war, the U.S. Navy introduced the Bat, a missile with autonomous radar guidance that successfully attacked ships and bridges at night at ranges of tens of kilometers, providing high levels of safety for the launch aircraft.

In the 1950s and early 1960s, however, interest in air-to-surface missiles (ASMs) was limited owing to concerns about cost, complexity, and reliability. The most widely used weapon was the U.S. Bullpup, rocket-propelled and guided by radio command. Several types of nuclear-armed ASMs were introduced, however, to permit strategic attacks to penetrate heavy air defenses. Due to the large effective radius of the nuclear warhead these missiles could employ preset inertial guidance, and some had ranges of more than 1000 kilometers.

The Vietnam War in the 1960s and early 1970s combined with advances in seeking and guidance technology to stimulate development of many new weapons. Again the weapons seeing widest use were American: Walleye and Maverick (TV guidance), Shrike and Standard ARM (homing on enemy radar emissions), and Paveway (homing on reflection from illumination by laser—so-called laser-guided bomb or LGB). These weapons combined miss distances of the order of a few meters with standoffs of around 5 kilometers or more. ASMs accounted for very few of the weapons employed in the Vietnam War, but achieved disproportionate results.

Armed helicopters intended for anti-tank attack were armed with adapted versions of ground anti-tank missiles. For anti-ship duties, specialized weapons were developed, of which the most widely employed has been the French Exocet, using autonomous radar homing. Similar to World War II's Bat in broad principle, Exocet was far more advanced, reliable, and effective. In the 1982 Falklands War, Argentine air forces destroyed or damaged a number of British ships with Exocets.

The Gulf War of 1991, in which a U.S.-led coalition rolled back Iraq's 1990 invasion of Kuwait, laid any remaining doubts concerning the value of modern ASMs. While many more conventional unguided bombs were dropped, ASMs (to include guided bombs, chiefly LGBs) achieved much the most critical and dramatic results. Iraqi radars, command posts, transportation and logistical facilities, bunkers, and tanks were picked off in great numbers with precisely guided ASMs. Iraqi ground forces, deprived of mobility and much vital matériel and quite demoralized, collapsed swiftly in the face of a powerful allied armored ground thrust, ensuring very low allied casualties.

Even in the U.S., the majority of air forces had not been equipped to make effective use of ASMs prior to the 1990s. The experience of the Gulf War led all who could afford to do so to rush to adopt ASMs. LGBs in particular, combining high precision with relatively low cost, quickly became all but universal. However, the Americans were not entirely satisfied with laser-guided weapons, which involved carriage of an expensive and bulky laser pod on the aircraft, only worked in reasonably clear weather, and required the aircraft to remain near the target throughout the LGB's fall. As the century closed, the U.S. introduced a new family of weapons that navigated to a prespecified point on the ground by reference to signals from the global positioning system (GPS) of satellites. The GPS-guided weapons were slightly less accurate than the LGBs (which remained in service) but more flexible. They require some means of determining the GPS coordinates of targets, implying a considerable increase in the technical sophistication of the collection, processing, and dissemination of intelligence about targets. Nevertheless, they were seen as an important step ahead and in a few engagements just at century's close performed well. Several variants were planned, ranging from two types of glide weapons allowing standoffs of around 50 kilometers (when released from high altitude) to powered weapons ranging hundreds of kilometers.

Hybrid weapons also were under development, employing GPS to bring them close enough to the target to permit transition to a autonomous precision homing system employing a laser radar, millimeter-wave radar, or passive imaging seeker. These gave promise of miss distances of a fraction of a meter against moving targets. Improvements in GPS were expected to cut miss distances of GPS-only guidance to around 1 meter.

As has been demonstrated fairly frequently, most guidance systems are vulnerable to some extent or another to countermeasures, whether technical or tactical. Many weapons have declined in effectiveness with use, as enemies took their measure. Nevertheless, by the close of the century it seemed that a powerful air force armed with sophisticated ASMs must enjoy a decisive advantage over an opponent not so favored. Thus, much more strongly than ever before, the ASM had transferred dominance in war from the side with

the biggest battalions to that with the most sophisticated air forces, a triumph of wealth and technology over mass.

WILLIAM D. O'NEIL

Further Reading

Dow, R.B. *Fundamentals of Advanced Missiles*. Wiley, New York, 1958. Many fundamentals remain relevant, in addition to its value in documenting earlier technologies.

Frieden, D.R., Ed. *Principles of Naval Weapons Systems*. Naval Institute Press, Annapolis, 1985. Extensive material relating to ASMs at modest level of technical complexity.

Gunston, B. *The Illustrated Encyclopedia of the World's Rockets and Missiles,* Crescent Books, New York, 1979. Capsule histories of developments from World War II through 1970s.

Ivanov, A. Radar guidance of missiles, in *Radar Handbook,* Skolnik, M.I., Ed., 2nd edn. McGraw-Hill, New York, 1990.

Lennox, D., Ed. *Jane's Air-Launched Weapons*. Jane's Information Group, Coulsdon, 2000. Earlier editions provide added information on earlier developments.

Pocock, R.F. *German Guided Missiles of the Second World War*. Ian Allan, London, 1967.

Zarchan, P. *Tactical and Strategic Missile Guidance*. AIAA, Washington D.C., 1990.

Missiles, Defensive

Missile defenses are complex systems composed of three major components: sensors to detect the launch of missiles and track them as they advance toward their targets, weapon systems to destroy the attacking missiles, and a command and control system that interconnects sensors and weapons. As a result of technological advances, these three components have evolved over the years since World War II, producing two major periods in the history of missile defense and suggesting the advent of a third by about 2025.

Between 1945 and the early 1980s, the Soviet Union and the U.S. developed and deployed missile defenses that used nuclear-tipped interceptors. The choice of nuclear weapons was dictated by technical limitations—interceptors of this period had to be guided from the ground using tracking information generated by ground-based radars. Given the limited accuracies that could be achieved with this approach to guidance, a nuclear warhead was required to achieve a reasonable probability of destroying an attacking warhead. However, there were drawbacks associated with nuclear-armed interceptors, including the stringent controls associated with nuclear weapons, the danger of stationing these interceptors near defended cities, and the fact that nuclear detonations would blind missile defense radars.

By late 1975, the U.S. and the Soviet Union had deployed operational missile defense systems. Under the 1972 ABM treaty and a 1974 protocol to the treaty, both systems were restricted to one hundred interceptors at a single site. While Russia continues to operate the Moscow site established by the Soviet Union, the U.S. closed its one Safeguard site at Grand Forks, North Dakota, in early 1976.

After deactivating Safeguard, the U.S. focused on research and development that would eliminate the need for defensive missiles with nuclear weapons. By this time, computers were becoming more powerful (operating faster with greater memory capacity) and decreasing in size. Phased array radars with electronic beam steering had continued to improve. Infrared sensors were becoming more and more sensitive, advancing from the simple one-detector devices of World War II to arrays with thousands of detectors. Furthermore, the outputs of these detectors were now being integrated and enhanced by the constantly improving computers.

Capitalizing on these advances, the U.S. Army developed and tested an experimental vehicle that destroyed its target through the kinetic energy generated when objects collide at high speeds. This hit-to-kill (HTK) interceptor was boosted outside the atmosphere where its own infrared sensor located the target and provided data to on-board computers that guided the interceptor to a collision with its target. In the early 1980s, the Army tested its new interceptor in a series of four flight tests known collectively as the Homing Overlay Experiment (HOE). After three failures, the test vehicle destroyed its target during the fourth test on 10 June 1984, demonstrating the feasibility of HTK interceptors.

As the HOE flights were taking place, the U.S. started the Strategic Defense Initiative (SDI) program. Focused on non-nuclear defense concepts, SDI started by surveying the missile defense technology base to see if advances since Safeguard would justify a new effort to develop missile defenses. Having determined that such an effort was justified, SDI leaders developed a new system architecture that included both ground- and space-based components that could attack and destroy a missile in all phases of its flight. At the same time, SDI produced significant advances in a broad array of missile defense technologies. As a result, virtually every missile defense component was made smaller, faster, and more powerful.

Illustrative of these developments are the following:

- The HOE kill vehicle of 1984 weighed 1100 kilograms. The EKV interceptor included in the midcourse defense system pursued under Presidents Clinton and George W. Bush performs the same mission but weighs only 55 kilograms.
- The infrared focal plane arrays used in surveillance satellites of the Defense Support Program during the mid-1980s, the most advanced IR sensors of that day, included 6000 IR detectors. Focal plane arrays developed under the SDI program included over 65,000 detectors, yet these devices were small enough to fit on the end of one's finger.
- The x-band radar that helped guide the EKV interceptor to the vicinity of its target was capable of ''seeing'' a golf ball at a range of over 3860 kilometers.
- The second stage engine on a 1980s Delta rocket weighed 45 kilograms and produced 4424 kilograms of thrust for a thrust-to-weigh ratio of 60 to 1. In one of its programs in the mid-1980s, the SDI program set out to develop motors that could produce 454 kilograms of thrust for each pound of engine. One result was the HEDI lateral thruster, which weighed 5.1 kilograms and produced 4800 pounds of thrust for a thrust-to-weight ratio of 930 to 1.

These advances fed into the missile defense programs pursued under the administrations of Presidents George H. W. Bush, William J. Clinton, and George W. Bush. They included progress in directed energy weapons (DEW) that promised to spawn a third phase in missile defense history, the DEW era. The applicability of lasers to missile defense had been recognized in the early 1960s, shortly after the laser's invention. By the 1970s a number of laser programs were under development within the U.S. Defense Department, and these programs continued into the 1980s when they were consolidated and expanded under the SDI program. Also included here was work on particle beam systems.

DEW systems possess two advantages over interceptor missiles. First and foremost is the velocity of their ''projectiles,'' bursts of energy, which are propagated at the speed of light. Second, directed energy weapons can be fired repeatedly. These two qualities allow these weapons to destroy multiple targets over long ranges.

At the beginning of the twenty-first century, the U.S. was developing the airborne laser (ABL) system that would mount a chemical oxygen–iodine laser in a modified Boeing 747-400 freighter aircraft. This system was expected to achieve initial operational status as early as 2009, when it would offer the ability to destroy ballistic missiles during their boost phase before they can deploy their multiple warheads and release decoys to confuse defenses. The ABL and other developments on the horizon suggest a third period in missile defense history, the DEW era, which could begin as early as 2025.

See also **Radar Aboard Aircraft; Radar, Defensive Systems in World War II; Radar, High Frequency and High Power; Radar, Long Range Early Warning Systems**

DONALD R. BAUCOM

Further Reading

Baucom, D.R. *The Origins of SDI: 1944–1983*. University Press of Kansas, Lawrence, KA, 1992.
Bruce-Briggs, B. *The Shield of Faith: A Chronicle of Strategic Defense from Zeppelins to Star Wars*. Simon & Schuster, New York, 1988.
Graham, B. *Hit-to-Kill: The New Battle over Shielding America from Missile Attack*. Public Affairs, New York, 2001.

Missiles, Long Range and Ballistic

During the 1960s, the U.S. and the Soviet Union began to develop and deploy long-range ballistic missiles, both intercontinental ballistic missiles (ICBMs) and intermediate-range ballistic missiles (IRBMs). The former would have ranges over 8000 kilometers, and the latter would be limited to about 2400 kilometers. The German V-2 rocket built during World War II represented a short- or medium-range ballistic missile. The efficiency and long range of these missiles derived from the fact that they required fuel only to be launched up through the atmosphere and directed towards the target. They used virtually no fuel traveling through near outer space. They were ''ballistic'' rather than guided in that they fell at their target after a ballistic arc, like a bullet.

Over the period of the Cold War, the U.S. Air Force deployed more than ten different ICBMs, considering the various models and modifications of the missiles. Sometimes the same warhead would be employed on different missiles, leading to some confusion or conflict among sources. The U.S. Air

Force ICBMs, in order of introduction, were as follows.

Thor: 1957–1975
Atlas: 1950s–1975
Jupiter C: 1950s–1960s
Titan I: 1950s–1960s
Minuteman I: 1962–1969
Titan II: 1963–1987
Minuteman II: 1965–time of writing
Minuteman III: 1970–time of writing
Peacekeeper: 1981–time of writing
Midgetman: cancelled

The Thor, Atlas, Jupiter C, and Titan missiles were liquid-fueled, presenting serious hazards for handlers and requiring considerable advance notice for fueling time. The Minuteman and Peacekeeper missiles, by contrast, were solid-fueled, safer, and could be kept on alert at all times. The Titan II, although liquid-fueled, could be maintained in a ready state. The explosion of a Titan II missile in its silo near Damascus, Arkansas in 1980 demonstrated the hazards associated with that type of missile and contributed to the decision to retire it.

By the mid-1980s, both the U.S. and the Soviet Union each had over 1000 such missiles, some with multiple independently targetable re-entry vehicles (MIRVs) on them. A 1985 estimate indicated that the U.S. had 2130 warheads on ICBMs, while the Soviet Union had 6420 warheads on similar missiles. At the peak of the Cold War in 1985, before reductions under arms-control agreements, the balance of long-range ICBMs was as shown in Table 1.

The Soviet SS20 was an intermediate-range nuclear missile introduced in 1978 that could threaten targets all across Western Europe. The U.S. responded by the deployment to Europe of Pershing II (an intermediate-range, or IRBM) and ground-launched cruise missiles (GLCMs) in 1983–1984 that could reach targets well inside the Soviet Union. These ‘‘Euromissiles,’’ although they could hold at risk strategic targets, were regarded by the U.S. as long-range theater nuclear forces, rather than as strategic weapons. With ranges of 1600 to 2400 kilometers, they were not intercontinental in range, but since they could target facilities within Soviet borders, they altered the Soviet’s perception of the nuclear balance of power. The Pershing II

Table 1 The balance of long-range ICBMs in 1985, before reductions under arms-control agreements.

	Number of launchers	Warheads per missile	Total warheads
U.S.			
Titan II	30	1	30
Minuteman II	450	1	450
Minuteman III Mark 12	250	3	750
Mark 12a	300	3	900
Totals	1030		2130
U.S.S.R.			
SS11	520	1	520
SS13	60	1	60
SS17	150	4	600
SS18	308	10	3080
SS19	360	6	2160
Totals	1398		6420

and the SS20 were sometimes called medium-range ballistic missiles (MRBMs) to distinguish them from ICBMs.

The Soviets argued that the Pershing II missiles and the Europe-based GLCMs should be considered in any treaty limiting the deployment of ICBMs. This issue, among others, became a stumbling block in negotiations toward SALT II, the second strategic arms limitation treaty. The issue was finally resolved by treating it under a separate intermediate-range nuclear force (or INF) treaty, signed in 1988, under which the GLCMs and the Pershing II missiles were removed from Europe.

On July 31, 1991, Presidents Mikhail Gorbachev of the Soviet Union and George H.W. Bush of the U.S. signed in Moscow the strategic arms reduction treaty (START I). The treaty reflected the trend in improved U.S.–Soviet relations that had been built over the prior five years. When President Ronald Reagan first proposed deep cuts in nuclear arsenals in May 1982, the concept had been called SALT III, but negotiations towards the eventual strategic arms reduction treaty did not begin in a serious fashion until 1990–1991.

The START agreement imposed equal ceilings or totals on megatonnage and on warheads on each side. The treaty also set up a complicated list of sublimits on the total number of delivery vehicles, including intercontinental ballistic missiles (ICBMs), submarine-launched ballistic missiles (SLBMs), and heavy bombers. The reductions were carried out in three phases over a period of seven years after the treaty came into force. After that seven-year period of implementation, each country would be allowed 1600 strategic nuclear delivery vehicles and no more than 6000 accountable warheads. Even though the treaty was not ratified by the Soviet Union before its dissolution in December 1991, Russia and the other three republics of the former Soviet Union then holding nuclear arms agreed to the terms of the agreement. Russia confirmed its adherence to START I in several legal steps, confirmed in agreement between President Bush and President Boris Yeltsin on June 16, 1992. The START agreement allowed each country to make up its allowed total of 6000 warheads by different combinations of weapons carried by bombers, ICBMs, and SLBMs, an arrangement that allowed the U.S. to take advantage of her long-range aircraft, and the Soviet Union to rely on her heavy ICBMs.

Both countries soon adopted a series of confidence-building measures to increase transparency and to provide early experience with verification techniques. For example, both countries opened ICBM silo hatches and displayed submarine missiles so that they could be counted by satellite. Later, under START II, both nations would witness on the ground the actual destruction of launchers, missiles, and other delivery systems.

The development of intermediate-range and tactical missiles by North Korea, Pakistan, India, and China and the modification of theater missiles by Iraq (such as the Soviet "Scud" or SS-1), all capable of carrying weapons of mass destruction (chemical, biological or nuclear warheads) continue to present threats to international stability in the twenty-first century.

RODNEY P. CARLISLE

Further Reading

Berman, R.P. and Baker, J.C. *Soviet Strategic Forces: Requirements and Responses*. Brookings Institution, Washington D.C., 1982.

Collins, J.M. *U.S.–Soviet Military Balance, 1980–1985*. Pergamon-Brassey's, Washington D.C., 1985.

Garthoff, R. *Détente and Confrontation: American Soviet Relations from Nixon to Reagan*. Brookings Institution, Washington D.C., 1994.

Gottemoeller, R. *Strategic Arms Control in the Post-Start Era*. Brassey's, London, 1992.

Pretty, R. *Jane's Pocket Book of Missiles*. Macdonald and Jane's Publishers, London, 1978.

Soviet Military Power. U.S. Government Printing Office, Washington D.C., 1981.

Tammen, R.I. *MIRV and the Arms Race*. Praeger, New York, 1973.

Tsipis, K. Arsenal: Understanding Weapons in the Nuclear Age. Simon & Schuster, New York, 1983.

Missiles, Long Range and Cruise

A cruise missile is an air-breathing missile that can carry a high-explosive warhead or a weapon of mass destruction such as a nuclear warhead for an intermediate range of up to several hundred kilometers. When launched from the ground, such missiles are known as ground-launched cruise missiles (GLCMs). Some historians of weapons technology regard the German V-1 or "buzz-bomb" operated in World War II, propelled with a ram-jet, air-breathing engine, as the first GLCM. The weapons do not require remote guidance, but automatically home in on pre-assigned targets, acting autonomously.

Modern American cruise missile development had its origin in a $20 million program initiated in the 1973 budget. Reliance on long-range ballistic missiles during the 1960s had led to the retirement of earlier cruise-missile models such as the

Regulus, Matador, Mace, and Hound Dog. From the 1970s onward, cruise missile development followed two paths, one for air-launched weapons (by the Air Force) and the other for ship-launched weapons for the Navy.

When modern cruise missiles were first under development in the 1970s and 1980s, there was considerable uncertainty about their circular error probable (CEP), with estimates predicted to run from under 30 meters to over 180 meters. The circular error probable is a measure of accuracy referring to a radius from the target point in which 50 percent or more of the missiles would strike. Thus a CEP of 30 meters represented a 60-meter circle with the target at its center in which one half or more of the missiles would strike.

Such a weapon, carrying a nuclear device, would be far more accurate than an intercontinental submarine-launched ballistic missile or than most intercontinental ballistic missiles (ICBM). A smaller nuclear device could be scheduled for a particular target if carried on a low-CEP GLCM. In fact, the development of low-CEP missiles and more powerful high explosives made it quite possible to develop non-nuclear war-fighting strategies that would be as effective in destroying enemy targets as earlier nuclear strategies. Accurate cruise missiles had the added advantage of reducing collateral (unintended) damage.

Among the best-known types of modern cruise missiles are the French-built air-launched cruise missile (ALCM), the Exocet; the American-built ship-launched cruise missile (SLCM), the Tomahawk; and the Chinese-built anti-ship ALCM, the Silkworm. GLCMs are distinguished from intermediate-range ballistic missiles (IRBMs) and ICBMs in that the cruise missiles have an internal system of targeting, and are air-breathing, while IRBMs and ICBMs are fired in a ballistic arc, utilizing either a solid or liquid rocket fuel together with oxygen to sustain the rocket at extremely high altitude and for exo-atmospheric flight. Since cruise missiles fly low and are air-breathing, they can be built lighter than IRBMs, which must carry an oxidizer, taking up weight that in a cruise missile can be devoted to warhead. The issue of the range of GLCMs led to controversies over how to classify them in arms control discussions between the Soviets and the Americans.

Following a policy announced in 1979, 464 American-built cruise missiles were stationed in Germany, Britain, Belgium, Holland, and Italy, in the period 1983–1985. Some of these highly accurate and long-range GLCMs, could reach targets deep in the Soviet Union. The press and weapons commentators called the GLCMs "Euromissiles." When the Soviets discussed strategic arms limitation or reduction in the mid and late 1980s, they wished to include the GLCMs, since they could reach strategic targets within their own borders. However, the U.S. insisted that the GLCMs were long-range theater nuclear force weapons, and should be covered in the treaty dealing with intermediate-range nuclear forces (the INF treaty) rather than in a strategic arms reduction treaty. The issue was difficult to resolve, but the GLCMs were finally included in the INF treaty negotiated in 1987 and signed in 1988. The GLCMs, together with 108 intermediate-range Pershing II ballistic missiles, were removed from Europe under that treaty.

Since cruise missiles are capable of carrying nuclear weapons, experts in nuclear proliferation include in their evaluations of nuclear-armed or nuclear proliferating states, such as Israel, Iraq, India or Pakistan, the weapons delivery systems. Domestically constructed cruise missiles or imported ones from suppliers in France, Russia, China, or North Korea add to the danger to neighbors represented by such nations. For example, China has provided Silkworm missiles and SLCMs to Iran.

The American Tomahawk land-attack missile (TLAM) was originally designed to be able to fit in a submarine's torpedo tube. In effect, by arming TLAM missiles with nuclear warheads, attack submarines could be converted into nuclear missile-launching platforms. The Tomahawk anti-ship missile (TASM) is another type of cruise missile. Since such a missile would carry 450 kilograms of high explosive, it would take only one such missile to disable a major warship. The TASM had a much longer reach than a torpedo.

Since cruise missiles can be launched from aircraft or ships many miles from their targets, they are classed as "stand-off" weapons. The launching platform, either ship or airplane, need not come within range of the defenses of the target. With the development in the 1980s of terrain-following and then laser-guided air-launched cruise missiles (ALCMs), extremely accurate missiles could be aimed at a particular airshaft, window, or door of a target building. Terrain-contour following missiles require that an accurate map of the ground over which the missile flies be loaded into an onboard computer. Such maps can be downloaded from satellite images. Cruise missiles have been employed with high-explosive warheads in several wars since the early 1980s. Particularly effective was the use by the Argentines of the

French Exocet during the Argentine–British war over the Malvinas (Falkland) Islands in 1982, and the use by Americans of ALCMs and SLCMs in operations over Iraq in 1991, Kosovo in 1999, and Afghanistan in2001.

RODNEY P. CARLISLE

Further Reading

Betts, R.K. Ed. *Cruise Missiles: Technology, Strategy, and Politics*. Brookings Institution, Washington D.C., 1981.
Cockburn, A. The air-launched version of the cruise missile: building on a foundation of failure? *Defense Week,* 23 June 1980.
Gray, C.S. *The Future of Land-Based Missile Forces*. Adelphi Paper no. 140, International Institute for Strategic Studies, London, 1978.
Huisken, R. The Origin of the Strategic Cruise Missile. Praeger, New York, 1981.
Jones, R. et al. *Tracking Nuclear Proliferation*. Carnegie Endowment, Washington D.C., 1998.
Tsipis, K. Cruise missiles. *Scientific American*, February 1977.

Missiles, Short Range and Guided

Early development of U.S. guided missiles can be seen as a direct case of technology transfer from the German V-2 program to the U.S. The Redstone missile, operational in 1956, was developed under the leadership of Werner von Braun and other German scientists who emigrated to the U.S. immediately after World War II. The project had initially been named "URSA," changed to "Major," and finally "Redstone." The guidance system was developed by the U.S. Army Ballistic Missile Agency at Redstone Arsenal, Huntsville, Alabama. The rocket motors were built by North American Aviation. The liquid-fueled rocket burned alcohol and liquid alcohol, and had a maximum range of the order of 280 to 320 kilometers, although later multistage missiles with the Redstone as the first stage reached ranges up to 4800 kilometers.

The upper limit of an intermediate range missile was arbitrarily set in the U.S. at about 2400 kilometers, and the upper limit of short-range missiles varied in official literature from 950 to 1300 kilometers. Other early short- and medium-range missiles developed by the U.S. are listed in Table 2.

In November 1956, U.S. Secretary of Defense Charles Wilson announced that the Army would have jurisdiction over ground-launched guided missiles with ranges under 320 kilometers, and that the Air Force would have control over those with higher ranges. Accordingly, the Army withdrew from the Jupiter program, and supported the Vanguard program to launch Explorer satellites. The Jupiter-C and prior work on Redstone missiles became part of that program. Juno I carried the first successful U.S. satellite into orbit on January 31, 1958.

One of the most difficult issues in early intermediate- and short-range missile work was perfecting a guidance system. Although ballistic missiles

Table 2 Early short- and medium range missiles developed by the U.S.

Missile	Service	Year	Manufacturer	Guidance Type	Range: km
Corporal	Army	1951	Firestone/Gilfillan	Radar	120
Sergeant	Army	1958	Sperry/Thiokol	Unguided	Artillery
Lance	Army	1972	Vought	Light beam, inertial	120
Honest John	Army	1953	Douglas/Emerson	Unguided	140
Jupiter	Army	1955	Chrysler/North American	Radar	
Matador	Air Force	1954	Martin	Ground controlled	900
Mace	Air Force	1958	Martin	Self-contained	900
Regulus I	Naval Bu Aer	1950	Chance Vought	Sperry intertial	800
Regulus II	Naval Bu Aer	1958	Chance Vought	Self-contained	1600
Thor	Air Force	1958	Douglas/North American	Inertial	2400

are generally regarded as unguided, of course they were guided in their initial stage; that is, pointed at the target as one would point artillery. True guided missiles, like cruise missiles, rely on systems that control the flight from launch to target, whereas ballistic missiles, whether long- or medium-range, were usually only guided or aimed during the initial lift-off and flight-path selection phases. Some confusion in the popular literature derived from this fact, and many ballistic missiles, guided only in lift-off, were known as guided missiles.

Early guidance systems for this initial stage were usually inertial, self-contained guidance systems. Inertial guidance systems relied on Newton's Second Law of Motion, and incorporated a computer (fairly primitive in early systems) to integrate velocity and distance with direction to target and accordingly control thruster angles and stability of the missile during ascent. Other systems included celestial guidance systems that corrected the basic inertial guidance with supplementary position and velocity information gathered from celestial star tracking, either optical or radio. A beam-rider system utilized a radar or light beam that the missile would follow in initial stages.

Short-range or artillery-type rockets surface-to-surface, or SS, missiles such as the Honest John (with ranges up to 37 kilometers) were simply aimed at the target. Mounted on a mobile transporter-erector-launcher (TEL), or on a truck, the vehicle would be pointed in the right direction, the launch rack elevated to the proper angle, and the missile simply launched in a parabolic arc. Such surface-to-surface missiles were guided by the aiming and elevation of the launching rack. The Lance, with a range to 120 kilometers, had a simple inertial guidance system, as did the 140-kilometer range Sargeant. The Pershing I, with a range up to 740 kilometers, had a more sophisticated inertial guidance system.

Short- and medium-range systems developed before the 1970s by other countries with a nuclear capability are listed in Table 3.

The most famous of the intermediate range missiles listed above is the SS-1 "Scud." Manufactured in several models with increasing range, they were sold abroad and used extensively by Iraq in the 1990–1991 Gulf War. Scud A had a range of 130 kilometers, Scud B a range 270 kilometers, and Scud C up to 450 kilometers. The Iraqis modified the Scuds further, by linking two together and achieving a range up to a reputed 900 kilometers, striking targets in Israel and deep in Saudi Arabia. In addition to initial aiming with the TEL, the Scud had a system of inertial guidance with fixed external vanes and movable auxiliary vanes positioned in the motor thrust efflux.

With most nations surrounding their military capabilities with considerable secrecy, different published range figures and guidance types sometimes contradicted each other. Furthermore, the distinction between an intermediate-range missile (up to about 2400 kilometers) and an intercontinental- or long-range missile was a matter of definition over which there was never complete agreement. Some publications would include submarine-launched missiles up to intermediate range in short- and medium-range ballistic missile listings. Although the missiles listed here were generally capable of carrying a nuclear warhead, most were also loaded with conventional explosive warheads.

RODNEY P. CARLISLE

Table 3 Short- and medium-range systems developed before the 1970s.

Country	Missile	Range: km	Guidance
France	Pluton	120	inertial
France	SSBS	3000	inertial, gimballed thrust motors
USSR	Scud (SS-1)	130–450	inertial
USSR	Frog	32	spin stabilized
USSR	Scaleboard	800	inertial
USSR	Sandal (SS-4)	1770	radio command and inertial
China	CSS-1	NA	inertial

NA: not available.

Further Reading

Collins, J.M. *U.S.-Soviet Military Balance, 1980–1985*. Pergamon-Brassey's, Washington D.C., 1985.

Merrill, G., Ed. *Dictionary of Guided Missiles and Space Flight*. Van Nostrand, New York, 1959.

Pretty, R. *Jane's Pocket Book of Missiles*. Macdonald and Jane's Publishers, London, 1978.

Soviet Military Power. U.S. Government Printing Office, Washington D.C., 1981.

Tsipis, K. *Arsenal: Understanding Weapons in the Nuclear Age*. Simon & Schuster, New York, 1983.

Walker, P.W. Precision guided weapons, *Scientific American*, August 1981.

Missiles, Surface-to-Air and Anti-Ballistic Missiles

In World War II, when Japanese kamikaze aircraft showed the amount of damage that could be inflicted with a single explosive-laden plane, it became apparent that machine gun and anti-aircraft fire were insufficient protection against current and future weapons. The answer was to combine radar detection, guided rockets, and the proximity fuse into surface-to-air missiles or SAMs. Intensive development in the postwar years produced the Sea Sparrow in the 1950s as one of the first, successful SAMS. When identifying a Warsaw Pact weapon as a surface-to-air missile, NATO forces would assign it an "SA" or surface–air, number.

Propelled by solid fuel-rockets, with radar guidance and or heat-seeking systems, and proximity fuses, there have been dozens of models and modifications of SAMs developed over the last half of the twentieth century. They have ranged from small shoulder-mounted devices effective against helicopters and low-flying aircraft to ship- and ground-based missile systems capable of destroying extremely high-flying aircraft or even ballistic missiles. In addition to SAMs that were loaded with high explosive, both the U.S. and the Soviet Union began to develop missiles, armed with nuclear warheads, which would be capable of destroying incoming intercontinental ballistic missiles (ICBMs). Such anti-ballistic missiles (ABMs) would be set to detonate outside the atmosphere, destroying incoming missiles with intense radiation or electromagnetic pulse. The distinction between a high explosive-loaded SAM with high-altitude capability and a nuclear-armed ABM was not always made clear in the press, nor in strategic literature.

The development of nuclear-armed ABM systems proceeded in the U.S. in the late 1960s, resulting in the creation of two missiles: Sprint and Spartan. Under the anti-ballistic missile agreement signed at the same time as the strategic arms limitation treaty (SALT-I) in 1974, both the U.S. and the Soviet Union agreed not to deploy ABMs at more than two sites in each country, later reduced to one site each. In the case of the U.S., a Spartan defense was built near ICBM silos at Grand Forks, North Dakota. However, the system was installed, tested, and then turned off. The Sprint and Spartan missiles were deactivated and stored.

The ABM treaty (which the U.S. renounced in 2002) prohibited ABMs capable of intercepting ICBMs, but it did not prohibit the development of surface-to-air missiles capable of destroying tactical or medium-range ballistic missiles, armed either with nuclear or high-explosive warheads. The U.S., the Soviet Union, and Britain proceeded to develop surface-to-air missiles with extremely high altitude capabilities. These SAMs that could serve as part of a ballistic missile defense (BMD) systems have sometimes been regarded as types of ABMs.

The Nike-Hercules, although intended as an anti-aircraft missile, could be loaded with either high explosive or nuclear warheads, and proved its efficacy against some ballistic missiles, when tested against a Corporal ballistic missile and against another Nike-Hercules. The announced altitude limit of the Nike-Hercules was 46,000 meters. Nike-Hercules missiles were deployed to NATO nations in the 1970s. Since NATO targets could be reached by intermediate-range, rather than intercontinental-range ballistic missiles, such deployment of the Nike-Hercules as a BMD system did not represent a violation of the ABM treaty.

The Nike-Hercules was followed by the Patriot primarily designed as an anti-aircraft ground-to-air missile, loaded with a high-explosive warhead. Development of the Patriot, begun in 1972, was delayed in 1977 by a decision to replace the missile's onboard computer system with a digital, rather than an analog system. Test firings in 1984 proved out the computer system. The core of the Patriot system was the Raythen MPQ-53 radar. Whereas the earlier Nike-Hercules required four separate radar types, the Patriot had a single system for surveillance, target acquisition, tracking, ranging and range rate, as well as missile tracking, command guidance and target illumination. The Patriot was carried on a two-axle semitailer and then leveled with jacks, and propelled by a solid-fuel Thiokol rocket motor.

With an announced altitude limit of 24,000 meters, the Patriot was only moderately successful

as a defense against the Scud in the Gulf War. When the U.S. deployed the Patriot in 1987, and then used it as a defense against the Soviet-made surface-to-surface Scud missiles during the Gulf War, 1990–1991, the Patriot, properly classified as a SAM, was often loosely referred to as a form of ABM. When used as a close-in defense system, the detonation of the Patriot could cause ground casualties or damage since it simply blew apart the incoming missile, letting large debris, engines, and fuel tanks rain down at or near the intended target.

The Soviets developed in the early 1960s a surface-to-air missile they designated the S-200 Volga, referred to in the West as the SA-5 Gammon missile. Western observers assumed that it was the Gammon that brought down the U-2 spy plane over the Soviet Union on May 1, 1960. The SA 5B, deployed in 1970, had a nuclear warhead; the SA 5C, deployed in 1975, had the option of either a nuclear or high-explosive warhead. With a potential altitude of 29,000 meters, the Gammon was thought to be capable of destroying incoming ballistic missiles. Western observers were divided over whether the SA 5B and 5C represented violations of the ABM treaty.

The Soviet SA-12B Giant surface-to-air missile was also capable of extremely high altitudes, over 30,000 meters. Like the SA 5s, it appeared to be in violation of the ABM agreements when introduced in 1986.

Many other anti-aircraft missiles designed and built in the U.S., France, the Soviet Union, and the U.K. were capable of high-altitude performance and many could have been capable of intercepting some ballistic missiles. Sweden, Japan, Israel, Italy and other countries developed a wide variety of mobile SAMs with altitude ceilings up to 3000 meters—quite effective against combat aircraft.

RODNEY P. CARLISLE

Further Reading

Chant, C. *Air Defense Systems and Weapons: World AAA and SAM Systems in the 1990s*. Brassey's Defence Publishers, London, 1989.

Pretty, R. *Jane's Pocket Book of Missiles*. Macdonald and Jane's Publishers, London, 1978.

Mobile (Cell) Telephones

In the last two decades of the twentieth century, mobile or cell phones developed from a minority communication tool, characterized by its prevalence in the 1980s among young professionals, to a pervasive cultural object. In many developed countries, more than three quarters of the population owned a cell phone by the end of the century.

Cell phone technology is a highly evolved form of the personal radio systems used by truck drivers (citizens band, or CB, radio) and police forces in which receiver/transmitter units communicate with one another or a base antenna. Such systems work adequately over short distances with a low volume of traffic but cannot be expanded to cope with mass communication due to the limited space (bandwidth) available in the electromagnetic spectrum. Transmitting and receiving on one frequency, they allow for talking or listening but not both simultaneously.

For mobile radio systems to make the step up to effective telephony, a large number of two-way conversations needed to be accommodated, requiring a duplex channel (two separate frequencies, taking up double the bandwidth). In order to establish national mobile phone networks without limiting capacity or the range of travel of handsets, a number of technological improvements had to occur.

The major problem with radio-based communications is the limited bandwidth available to transmit all the data of the calls that people wish to make. The high-frequency region of the electromagnetic spectrum has little space left in which to transmit and receive calls. To efficiently use the allocated bandwidth, mobile phone operators began by splitting it up into smaller segments (channels). Before networks were established, a single, high-power base antenna would cover a large area, but within the radius of the transmitter (normally greater than 32 kilometers), the number of users was strictly limited by the small number of channels available within the bandwidth.

An early step toward establishing a network was introduced in St. Louis, Missouri, in 1946, with a further 25 American cities adopting similar services in the next year. These were the first "zoned" systems, with multiple receiving stations, allowing users to travel around the city transmitting calls from car-based units to the nearest available receiver (this was controlled automatically in a mobile telephone switching office (MTSO) based on the relative signal/noise ratios detected). However, all transmissions from the mobile radiotelephones were picked up by a central station, making the systems again limited by the small number of channels available. This lack of space also required calls to be "half-duplex" only, without simultaneous talking and listening.

Research by Bell labs in the 1940s and 1950s showed that channels could be reused across the

network if the areas using the same channels could be separated sufficiently to remove cochannel interference. This was a huge breakthrough, but it would still be three decades before mass-communication systems were up and running. During this time, transmitters and receivers became more efficient, so channels required less bandwidth and more could be squeezed into the same space. In the 1980s, governments began to allocate bands of the electromagnetic spectrum exclusively for nationwide mobile communication. These developments allowed for the implementation of modern cellular networks, which removed the long waiting lists caused by years of excess demand.

Cellular networks require small base stations to communicate on a duplex channel with the mobile phones within their cell (area of coverage). The system is modeled around a hexagonal honeycomb, with a base station at the center of each hexagon (see Figure 9). Each cell has a limited number of channels assigned to it, which can be reused without interference in nonadjacent cells. Crucially, as more users are added to each network, the number of channels can stay the same and the cells can be split into smaller cells by reducing the power and range of base stations. This allows for almost unlimited increases in capacity. As a phone travels from cell to cell, the MTSO allocates the phone to the antenna with the strongest signal, switching the frequency channel to continue the call without interruption. This is known as a "handoff." The phone can travel across the whole network by being regularly reassigned channels (i.e., frequencies).

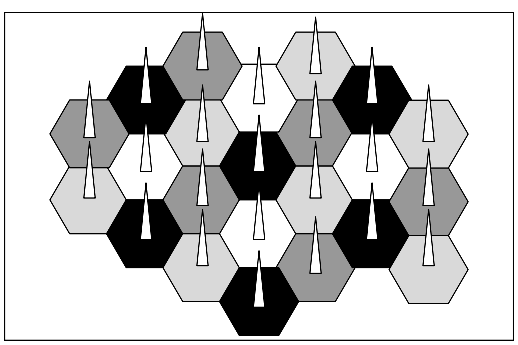


Figure 9. A representation of a network with a four-cell cluster arrangement, which allows four lots of channels to be reused in non-adjacent cells. In the diagram, the four shades represent cells that can re-use the same frequencies without interference if the system works perfectly. In reality, cells more closely resemble overlapping circles rather than hexagons, so a seven-cell cluster is normally used, increasing the distance between cells using the same channels and reducing the likelihood of interference.

The network system obviously demands a large number of base stations in busy areas, but these allow clear communication at low power, which in turn allows for smaller batteries, reducing the size and cost of the phones. In the 1980s, improved network coverage along with advances in battery efficiency, computer hardware (microchips and integrated circuits), and software allowed mobile phones to become truly mobile. What was once exclusively a business tool, requiring a cumbersome separate battery unit, could be easily slipped into a pocket and used relatively cheaply, often completely replacing a traditional phone line.

Mobile phone networks cover most urban areas of the world and continue to expand. With increased coverage and greatly reduced cost, the number of users increased more than ten-fold in the 1990s. In 2000 alone, around 400 million handsets were sold. It is estimated that by 2005, there will be more than 1 billion mobile phone subscribers worldwide.

In the early 1990s, there was a worldwide move to a digital standard for mobile phones. Digital compression techniques allow for a greater volume of information to be transmitted within the same frequency band, increasing the capacity of each cell. In Europe, GSM (global standard for mobile communications) operating on the 900 megahertz band, became the digital standard (although many countries, including the U.K., now have a separate 1800 megahertz band used by newer networks). In the U.S., mobile phone systems operate at 800 and 1900 megahertz on the PCS (personal communication system) standard. Although the systems are incompatible, both use a technique called time division multiple access (TDMA). This works by splitting the available transmission time on each channel into a number of time slots (typically eight), allowing more users to simultaneously use each channel.

Digital mobile telephony also allows for the reliable transmission of data. One development is the addition of WAP (wireless application protocol) facilities to handsets, which permits rudimentary Internet access. However, slow downloads and a paucity of available information have prevented WAP from being adopted as quickly as other less advanced features such as SMS (short message service) text-messaging, which has revolutionized the way many users interact with their phones. A third generation of mobile phone technology (the analog standard is the first and digital the second) promises to further change the way we use our handsets, offering multimedia and respectable Internet access, but this will involve

the construction of new networks in the twenty-first century.

See also **Telecommunications; Telephony, Digital**

JACK STILGOE

Further Reading

Akaiwa, Y. *Introduction to Digital Mobile Communication*. Wiley, New York, 1997.

Lee, W.C.Y., *Mobile Cellular Telecommunications: Analog and Digital Systems*, 2nd edn. McGraw-Hill, New York, 1995.

Lee, William C.Y. *Lee's Essentials of Wireless Communications*. McGraw-Hill, New York, 2000. A less technical book by the same author, introducing the relevant issues.

Lewis, W.D. Bell Telephone Laboratories, coordinated broadband mobile telephone system. *IRE Transactions*, May 1960, p. 43.

Garg, V. and Wilkes, J. *Wireless and Personal Communications Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1996.

Rahnema, M. Overview of the GSM system and protocol architecture. *IEEE Communications Magazine*, April 1993.

Schulte, H.J. Jr. and Cornell, W.A. Bell Telephone Laboratories, multi-area mobile telephone system. *IRE Transactions*, May 1960, p. 49.

Schwartz, M. *Telecommunication Networks*. Addison-Wesley, Reading, MA, 1987.

Modems, *see* **Electronic Communications**

Motorcycles

Without the bicycle there was nothing to motorize. Originating in France in the late 1770s, the hobbyhorse became the safety bicycle over 110 years through many people's efforts, including Thomas Humber, Dan Rudge, George Singer, James Starley, and Harry Lawson. Lawson's patent of 30 September 1879 shows a pedal-powered, rear-wheel chain-driven bicycle, something perfected by John Kemp Starley and William Sutton in 1885. The technology to motorize a bicycle had existed for years. A German-built steam-driven vélocipède is shown in a drawing dated 5 April 1818. Sylvester Roper of Roxbury, Massachusetts, exhibited a bicycle with a charcoal-fired two-cylinder engine at fairs in the eastern U.S. in 1867, and in France, the Michaux-Perreaux was built in 1868 by attaching a small commercial steam engine to a bicycle.

These attempts suffered from the disadvantageous power to weight ratio of steam engines. The solution came in 1883, when Gottlieb Daimler and Wilhelm Maybach perfected the high-speed internal combustion engine. Applications of the technology to motorize bicycles came quickly, and included work by two English pioneers: Edward Butler, who built his "Petrol-Cycle" in 1887; and J.D. Roots, who developed a motor-tricycle in 1892.

Brothers Henry and Wilhelm Hildebrand, Alois Wolfmuller and Hans Greisenhof made the first commercial motorcycle in Germany. They developed their machine from 1892 and manufactured it from 1894 as the Hildebrand and Wolfmuller "Motorrad," meaning motorcycle, the first use of the word. A number of Hildebrand & Wolfmullers were imported into Britain, where Colonel H. Capel Holden also produced motorcycles from 1897.

In 1895 the French DeDion–Buton Company built a small, light, high-revving four-stroke single engine that used battery- and coil-ignition. It was 138 cc, developed 0.5 horsepower, and made the mass production of motorcycles possible. DeDion-Buton used it in tricycles, but it was widely copied, including by Indian and Harley-Davidson in the U.S.

By 1900, at least 14 companies had tried their hand at motorcycle production in Britain. The twentieth century added to their number, and in all 639 makes of motorcycle would be produced in the U.K., the most of any country in the world. The first motorcycles retained the bicycle's "diamond" frame, into which the mechanical parts were fitted, the majority going between the main down tube and front fork, the engine driving the rear wheel through a belt. They had no gearbox or clutch, being fixed speed. Riders pushed the bike forward, ran alongside to start it, and then jumped aboard. Some manufacturers fitted clutches, which permitted stationary starting and allowed the engine to run without the bike moving and was favored by amateur mechanics.

Most early motorcycles had a four-stroke engine, usually with a single cylinder. This had inherent drawbacks. Only the third of the four strokes was powered, and on the second, or compression stroke, the piston was furthest from its last power stroke and encountering its greatest resistance. Without careful setting of valve timings, single-cylinder four-stroke engines could run erratically.

Two ways were found around these drawbacks. By 1905 V-twin engines had been produced in Europe and the U.S. These have two cylinders, usually angled at between 26 degrees and 30 degrees apart. By having one cylinder working two strokes ahead or behind the other, the engine runs on more power strokes, more smoothly, with

more power. The other way around the drawbacks was the two-stroke engine. Sir Dugald Clerk developed an engine working to a two-stroke principle in 1881. This used two-cylinders, one to compress the mixture and another in which it was exploded. Work to produce a single-cylinder two-stroke engine was undertaken by Alfred Scott between 1900 and 1908, which he used in his company's motorcycles from 1909; followed by Levis from 1910, and Velocette from 1912.

Sporting experience showed the value of gears. Some of the first were adjustable, like those on the Zenith Gradua in 1909, and the Rudge Multi in 1911. A rider-operated lever varied the diameter of an engine and rear wheel drive pulley, giving a variable ratio. More conventional two-speed gears were more widely fitted from around 1910, and by 1914 three- and four-speed gearboxes were fitted by some manufacturers.

Early motorcycles used belt-drive, the use of a v-section one being widely adopted. Belts were wont to stretch, slip, or break; but they were still used on single-gear, sports and lightweight machines into 1920s. Some manufacturers favored drive chains; especially on twin-cylinder models where the extra power could cause slippage. A compromise was "chain-cum-belt" drive. Here the drive from the engine to the gearbox was by chain, and that from the gearbox to the rear wheel by belt.

The British government suspended motorcycle production late in 1916 for aircraft and munitions work. It resumed by 1919. Many servicemen had ridden motorcycles and wanted to own one, providing a boom time for new makers. The 1920s also saw changes to motorcycle design and specification. Frames became elongated and low-built, the saddle was set further back, nearer the ground, and petrol tanks were mounted on top of the frame, in front of the saddle, rather than suspended beneath the top tube. Transmissions progressed too. Veloce introduced the first foot-operated gear change in 1925, the decade also seeing the progressive adoption of all-chain drive by most makers.

The marketing of motorcycles changed too. From around 1928 manufacturers named their machines. Thus Model 7s, in "Standard" or "De Luxe" versions, became a "Big Twin Export," "Speed Chief," or "Super Sports." Replicas of motorcycles successful in races and trials were also produced, as were machines for the imported Australian sport speedway.

With recession in the late 1920s, fewer motorcycle firms were formed, and established ones experienced difficulties. A number went under, while others had to cut costs and prices to survive. By the early 1930s serious and ultimately fatal damage had been done to the motorcycle industry in Britain and elsewhere. The 1930s was a decade of consolidation, characterized by the formation of Associated Motor Cycles Ltd (AMC), which eventually owned some of the greatest names such as Sunbeam and Norton.

European governments generally had a more enlightened view of motorcycles and their use, and through a combination of exemption from tax, based either on weight or engine capacity, or from a riding test, they fostered a market for lighter weight machines and a climate in which manufacturers looked for ways to make frames ever lighter and small engines ever more powerful.

Technical innovation came through the test bed of racing, where, in the 1930s, the German firms BMW and DKW, and the Italian Guzzis challenged British dominance. They produced lighter machines with higher revving engines, which, by 1936, enjoyed great success in competition. The threat this posed to the traditionally conservative British motorcycle industry was compounded by the fact that both of these countries had totalitarian governments.

World War II stifled motorcycle production, but the postwar period looked brighter. Prewar racing experience using sprung frames, introduced by Guzzi in 1935, and by Norton in 1936, influenced the design of road machines after 1945. Pressed steel frames were also increasingly common, offering both an economy in manufacture and weight, with the added advantage of better weatherproofing and enclosed engines. The use of new lightweight alloys allowed engine development to proceed apace, with single- and multicylindered versions appearing which offered power outputs that equaled or exceeded those of traditional designs, from units weighing considerably less and offering greater fuel economy.

The use of these new materials allowed manufacturers to meet the growing demand for personal motor transport seen throughout Europe and the Far East after the war. Two new designs of motorcycle appeared—the cyclemotor, or moped, and the motor scooter—and smaller motorcycles were remodeled as miniature or lightweight machines. Mopeds retained pedals and generally had small engines in the 49 to 98 cc range. One of the earliest was the NSU Quickly in Germany, but Japanese manufacturers soon joined the market, most notably Honda, who began makeshift production in October 1946. The ultimate development of the form came in July 1958 with the

introduction by Honda of its C series Super Cub machines. Known colloquially as the "Honda 50," its 30 million sales to date make it the world's best-selling motor vehicle.

The motor scooter was an Italian concept that owed much to the aircraft engineering experience of its manufacturer Piaggio & Co. Their Vespa design, introduced in April 1946, was an instant success, and was followed in 1948 by Innocenti's Lambretta. Imported Italian scooters proved popular, and some British firms introduced their own versions in response, but whereas the former had style and were lightweight, British scooters were heavy and far from handsome.

As motorcycle weights were reduced and their power increased, improved brakes were needed, with hydraulic systems replacing cables and then the introduction of disk brakes, first on the front wheel and then on the rear. Racing experience continued to influence the design of road machines, notably in the use of streamlining and fairings, especially on Italian motorcycles.

The Ariel Leader of 1958 represented the first major departure from traditional motorcycle design by a British manufacturer since the war. It used unit construction, with all its mechanicals enclosed behind steel panels, and power coming from a 247-cc two-cylinder, two-stroke engine mounted on to an integral crankcase and gearbox casting. Externally, fairings and a windshield offered a previously unheard of level of protection to the rider on a machine that was very handleable. Despite its innovation, at a shade over £200, the Leader was expensive; scooters were cheaper to buy and run, and £50 more would buy the more powerful Triumph 500 Twin. Competition was also on the horizon, and 1959 saw the launch of both the Honda 250 cc C72 Dream motorcycle, with a pressed steel frame, swing arm and front-leading link forks, sophisticated OHC all-aluminum engine, electric starter and indicators, and the Austin Seven and Morris Mini Minor motorcar, which offered a family four wheels for £400.

Honda led the onslaught on the British and U.S. markets by Japanese manufacturers. Interest in the machines was generated by their higher technical specification; electric starters were a standard on Hondas Yamahas and Suzukis by 1960. Japanese firms also built their reputation on racing success, and in 1961 Honda stunned the racing world with Mike Hailwood's twin victories at the Isle of Man, the first of an unprecedented string of wins. In the U.S. Honda made their impact through advertising, beginning in 1962 with their clever "You meet the nicest people on a Honda" campaign that directly addressed the myth that motorcycles were only for tough guys and rebels. It reached out and made Honda and motorcycling in general, appealing to everyone, also opening up a vast new market to other Japanese firms, such as Kawasaki, who began volume motorcycle production in 1962.

The 1960s saw the virtual collapse of the established motorcycle industry in Britain, especially after the failure of AMC in 1966. In contrast European and Japanese manufacturers responded to the new markets in the U.S. and Australia, where high-performance motorcycles continued to woo people out of cars and on to two wheels. A luxury high performance motorcycle market began in 1965 with the launch of the Honda CB450, which offered a 163 kilometers per hour top speed, and the Italian Guzzi V7. This set a trend that continued, and led to the development of ever more powerful motorcycles, and the coining of a new word—superbike—first applied to the Honda CB750F on its launch at the 1968 Tokyo Motorcycle Show. A 750-cc machine, with four cylinders and disk brakes, the CB750F was the biggest motorcycle out of Japan to date, and it proved that a high-performance motorcycle could also be very reliable.

See also **Automobiles, Internal Combustion**

PAUL COLLINS

Further Reading

Collins, P. *British Motorcycles since 1900*. Ian Allan, Shepperton, 1998. A technological and sociological analysis of the motorcycle, plus a survey of all 639 makes of British motorcycle.

Caunter, C.F. *Motor Cycles: A Historical Survey*. Her Majesty's Stationary Office, London, 1982. Exhaustive technological detail on the development of motorcycles.

Hough R. and Setright, L.J.K. *A History of the World's Motorcycles*. George Allen & Unwin, London, 1973. World view of motorcycle history.

Motorways, *see* **Highways**

N

Nanotechnology, Materials and Applications

The *Shorter Oxford English Dictionary* defines nanotechnology as "the branch of technology that deals with dimensions and tolerances of 0.1 to 100 nanometers." Thus nanotechnology is defined not by discipline but by size, and chemists, medical scientists, and electronic engineers working on this scale may all call themselves nanotechnologists. Nanotechnology is at the interface between physics, chemistry, engineering, and biology; the fundamental processes of living matter occur on the nanometer scale. A nanometer is 10^{-9} meters or about 4 atoms wide; a human hair is about 70,000 nanometers thick.

The impetus for nanotechnology came from a famous talk by the Nobel physicist Richard Feynman in 1959. Feynman observed that there seemed to be no natural lower limit of size that would constrain the design and manufacture of very small devices. His argument was pursued in the 1980s by the American Eric Drexler, who devised a conceptual framework for nanotechnology. He envisaged "assemblers," devices controlled by computer that would be able to manipulate individual molecules or even atoms, and which would be

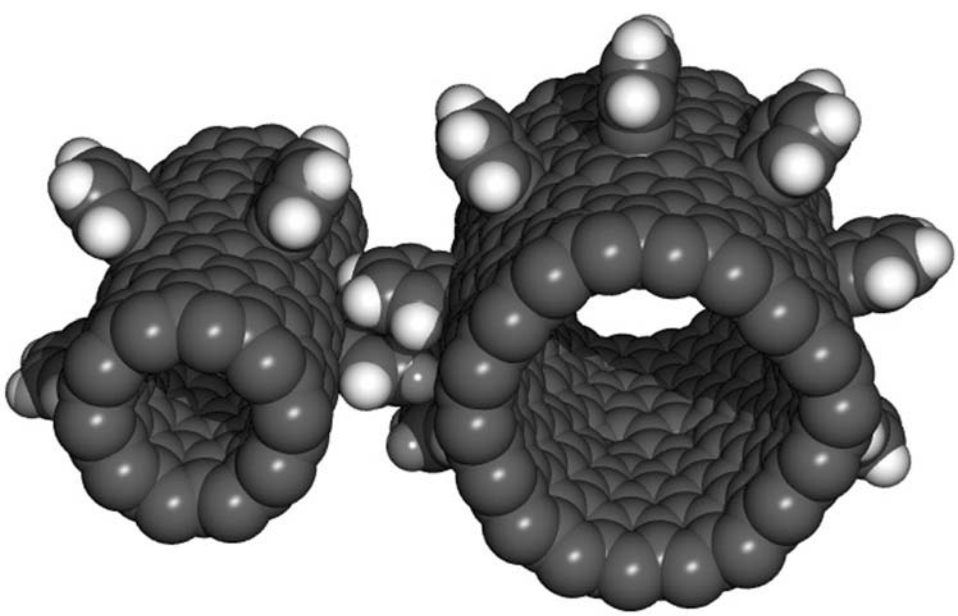


Figure 1. Simulation of fullerene gear train.
[*Source: http://science.nas.nasa.gov/Groups/Nanotechnology/gallery/*]

able to build replicas of themselves. Drexler made large claims for nanotechnology as an enabling technology that would change the human future. Its most exotic applications were to be in medicine: sensing systems and repairing machines would be implanted within individual cells; tiny robots would roam the bloodstream, seeking out and destroying viruses, delivering chemotherapeutic drugs directly to cancer cells, reaming out clogged arteries. Eventually self-repair would permit people to live forever. Such ideas initially attracted more attention from speculative writers than from the scientific community. However, technology has now begun to catch up with his visions, and some important milestones of nanotechnology have already been passed in the laboratory.

"Bottom-up" nanotechnology refers to the construction or assembly of devices or structures molecule by molecule from their constituent atoms, using techniques that have no conventional counterpart such as the scanning tunneling microscope (STM) and electron beam lithography. The STM, developed by Gerd Karl Binnig and Heinrich Rohrer in IBM's Zurich Research Laboratory in 1981, can actually pick up individual molecules on the tip of its probe and reposition them under control, demonstrated in 1990 when the American researcher Don Eigler was able to write the letters IBM on a nickel substrate with 35 individual xenon atoms. "Motors" have been constructed of less than 100 atoms in size, powered by biochemical reactions or by incident light. At the moment these have only a limited range of movement but have demonstrated the principle. The spherical molecule fullerene, C_{60}, which was discovered in 1985 and earned its discoverers the Nobel Prize in 1996, has a diameter of 1 nanometer and offers the possibility of molecular-scale ball bearings for small machines. In 1996 James Gimzewski at IBM Zurich manipulated fullerene molecules with an STM to build a 10×10 abacus with an overall diameter of less than 1 nanometer (see Figure 2).

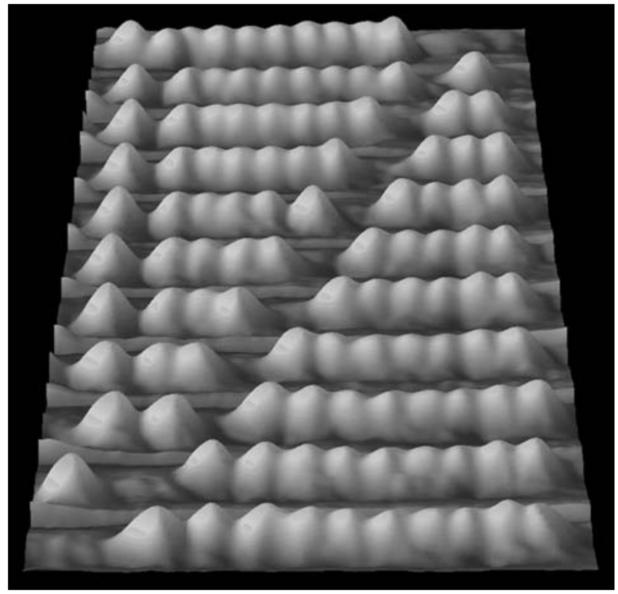


Figure 2. Abacus constructed from fullerene molecules. [*Source: IBM Zurich http://www.zurich.ibm.com/pub/hug/PR/Abacus/abacus.gif*]

Materials based on carbon chemistry appear to be the most promising building blocks for nanotechnology. Diamond was suggested by Drexler. In practice, rigid nanostructures have successfully been made out of DNA molecules, but currently the most popular building element for nanodevices is the carbon nanotube. Nanotubes, discovered in 1991 by Sumio Iijima of Japan, are single molecules of carbon in sheet form that can be rolled into tubes with diameters of a few nanometers but lengths of up to 1 millimeter. Nanotubes have already been used to make weighing devices and have successfully been rolled across a flat surface, suggesting their suitability as machine elements. Nanotube bearings appear to be wear-free. Nanotube circuit elements fabricated from the mid-1990s offer the possibility of molecular computing using single-molecule switches to process data. Carbon nanotubes have been commercially available since 1999, but cost is still a problem; weight-for-weight, nanotubes are more expensive than gold.

"Top-down" nanotechnology refers to the miniaturization of existing mechanical and electrical devices using modified forms of existing fabrication techniques such as etching or micromachining. Microelectromechanical systems or MEMS are already in use in a wide range of technical applications. In medicine, micrometerscale MEMS devices are available to control internal delivery of drugs. Microfluidic devices also exist capable of electrically and chemically interfacing with single cells. Other MEMS devices in production in the early twenty-first century include tiny accelerometers for activating airbags and optical sensing and projection chips mounting a million micromirrors. The hope is that at least some of these devices can be scaled down to nanosize NEMS, though a critic has drawn attention to the dangers of merely "reinventing the Swiss watch industry."

An everyday example of a MEMS device, part of which already operates in the nanoscale region,

is the read–write head on a computer hard drive. Although the active area of the head itself is some tens of micrometers in diameter, it "flies" above the disk platter on an air cushion that is only a few nanometers thick on the latest laptops. Because the operating clearance is smaller than the mean free path of the air molecules, the ideal gas laws can no longer predict the behavior of the air cushion, and the "wings" of the head must be elaborately contoured to stabilize its flight. So flat must the surfaces of platter and flier be to avoid moving contact that in static contact molecular forces will weld them together. Therefore special areas textured with nanoscale pimples must be provided on the platter to prevent intimate contact when the head is parked. The consequent metrological problems are likely to apply to other NEMS devices.

Nanotechnology has yet to make any significant impact on society, and has been attacked as a "cargo cult" pseudoscience that promises more than it can ever deliver. Even its supporters admit that medical robots small enough to operate in the bloodstream are at least a generation away. So far most of the progress has been top-down, but in the long term the bottom-up approach is likely to yield more important and fundamental results. However, several of Feynman's suggestions have already been realized, and large financial and technical resources are currently being invested in bringing these to market; the U.S. government earmarked $675 million for nanotechnology research in 2002. It seems likely that at least some of Drexler's predictions will become reality within a few years.

T. R. Thomas

Further Reading

Allegrini, M., Garcia, N., Marti, O., Eds. *Nanometer Scale Science and Technology*. IOS Press, Amsterdam, 2001.

Bhushan, B., Ed. *Handbook of Micro/Nanotribology*. CRC Press, Boca Raton, FL, 1998.

Drexler, E.K. *Engines of Creation*. Fourth Estate, London, 1996.

Fujimasa, I. *Micromachines*. Oxford University Press, Oxford, 1996.

Gross, M. *Travels to the Nanoworld: Miniature Machinery in Nature and Technology*. Perseus, Cambridge, MA, 2001.

Lyshevski, S.E. *Nano- and Micro-electromechanical Systems*. CRC Press, Boca Raton, FL, 2000.

Regis, E. *Nano: The Emerging Science of Nanotechnology*. Little, Brown & Company, London, 1996.

Stix, G. Waiting for breakthroughs. *Scientific American*, April 1996.

Timp, G.L., Ed. *Nanotechnology*. Springer Verlag, New York, 1998.

Navigation, *see* **Global Positioning System (GPS); Gyro-Compass and Inertial Guidance; Radionavigation**

Neurology

Following the neuroanatomical discoveries of Englishman Thomas Willis in the seventeenth century and Scotsman Alexander Monro "Secundus" in the eighteenth, and the neurophysiological work of Frenchmen Pierre Flourens, Guillaume Duchenne, and Jean-Martin Charcot in the nineteenth, knowledge of the brain and nervous system advanced significantly. But rapid progress in the clinical neurological sciences occurred only after discovery of x-rays by the German Wilhelm Roentgen in 1895 and subsequent developments made it possible to create images of living brain and nerve tissue. The diagnostic science of neurological imaging is broadly called neuroradiology.

In chronological order of development, six general neuroradiological techniques have dominated: plain x-ray films, pneumography, radiopaque myelography, cerebral angiography, computed tomography (CT), and magnetic resonance imaging (MRI). The development of the first four techniques is generally complete, but the fifth and sixth are still being refined. Other techniques of neurological examination include echoencephalography, electroencephalography, various ultrasound applications, and removal of cerebrospinal fluid (CSF) from subarachnoid space. The era from plain films to cerebral angiography took roughly 75 years, from 1898 to 1973.

Plain X-Rays

Plain x-ray films, available after 1896, were first used extensively for skull pictures in the Spanish–American War. Austrian Artur Schüller, the acknowledged father of neuroradiology, investigated their application to intracranial diagnosis and published his groundbreaking work in 1912. Swede Erik Lysholm refined the photographic method in the 1920s and 1930s. His manual of "skull tables" allowed precise identification of intracranial anatomical relations. Even as recently as the 1960s, physicians preferred plain films for some kinds of neurological diagnosis.

Pneumography/ventriculography

Pneumography, invented by American Walter Dandy in 1918, is the x-ray photography of the

skull (pneumoencephalography and pneumoventriculography) or spine (pneumomyelography) after air has been introduced. Dandy observed that abdominal surgical x-rays showed gas as black against the gray of soft tissues and the white of bones, and he had read American William Henry Luckett's surgical case report of air from the sinuses entering a fractured skull and showing the ventricles and subarachnoid space in an x-ray. He discovered that if air was introduced through a lumbar puncture needle, a pneumoencephalogram could be produced. His pneumographic procedure was dangerous in cases of brain tumors or papilledema because the increased pressure could herniate the brain stem. For such cases, making holes in the skull and putting the needles directly into the ventricles (ventriculography) was safer.

Formaldehyde, first used as an anatomical fixative in 1893, enabled the pioneer work of Swiss psychiatrist Adolf Meyer in the postmortem confirmation of neuroradiological diagnoses. Meyer discovered that the precise anatomical relations of the brain can be preserved for inspection if formaldehyde is injected into the CSF and allowed to stand 24 hours before autopsy. By this method, Meyer learned that what often killed brain cancer patients was not the tumors themselves but the herniation of brain tissue through the tentorial notch or into the foramen magnum. Meyer described sideward and downward shifts of the brain tissue in 1920; American Arthur Ecker described upward shifts in 1948. Transtentorial herniation became a common topic of neurosurgical research in the 1950s.

By 1934 Dandy had refined his diagnostic procedure into a standard routine. His patients at Johns Hopkins University would first undergo a general and neurological exam by a surgical resident, a neuro-ophthalmological exam by Frank B. Walsh, and an otological exam by Benjamin M. Volk. The next day Dandy himself would perform ventriculography and make his judgment on the basis of all these reports. Errors were often made, and the mortality rate was high, even for benign intracranial tumors; but at the time this method was the best available in America.

Ventriculographic technique was refined to a higher level in Sweden in the 1930s, primarily by Lysholm's team directed by Herbert Olivecrona. Pneumography, especially pneumoencephalography, was improved from the 1920s through the 1950s by Germans Otfrid Foerster and Erich Fischer-Brügge, Americans Leo M. Davidoff and Cornelius G. Dyke, Briton Henry Head, Australian E. Graeme Robertson, and Italian Giovanni Ruggiero.

Radiopaque myelography

Radiopaque myelography, x-rays of the spine with a variety of contrast media, was begun in 1921 by Frenchmen Jean Athanase Sicard and Jacques Forestier. Since then, the quest for safer, more effective, more easily applied contrast media is a major part of the history of neuroradiology.

Cerebral angiography

In 1926 or 1927 Portuguese physician and statesman Antonio Caetano de Egas Moniz invented cerebral angiography, x-rays of the skull after introducing a contrast medium into both carotid arteries. Egas Moniz mounted his camera on the "radio-carousel" invented by his colleague José Pereira Caldas to get a large series of angiograms in rapid succession. His earliest contrast media for cerebral angiography were very dangerous substances such as strontium iodide. A colloidal thorium dioxide solution, Thorotrast, was com-

Figure 3. Cerebral angiogram made by Arthur D. Ecker, M.D., in 1949.
[*Courtesy of the Department of Historical Collections, Health Sciences Library, SUNY Upstate Medical University*.]

monly used after 1929 because it provided better contrast and was somewhat safer.

Egas Moniz's method discouraged patients and physicians alike because it required incisions to expose both carotid arteries and left unsightly scars on the neck. The development in 1936 of percutaneous carotid injections of Thorotrast alleviated that difficulty. But Thorotrast is radioactive, and if any of the injected solution fell outside the carotid artery it would cause proliferation of neck tissue with disastrous results. In the 1940s some organic iodides, such as Perabrodil, were found to be safer than Thorotrast. Yet even into the 1950s some physicians argued in favor of Thorotrast and against percutaneous carotid injections.

The U.S. lagged behind Europe, especially Sweden, in the acceptance and development of cerebral angiography. In 1951 Ecker, who had translated Egas Moniz's method for his own use in the 1930s, published the first American monograph on the topic. The Anglophone world to its disadvantage neglected the German work of Fischer-Brügge and Hermann Coenen in the early 1950s.

Computed tomography

Mathematical equations may sit idle long after their discovery until someone finds practical uses for them. Such was the case with the equations necessary for the development of computed tomography (CT) scans, the greatest single advance in radiology since x-rays. Austrian Johann Radon discovered equations for determining plane functions from line integrals in 1917, but South African physicist Allan Macleod Cormack only learned of them in 1972, nine years after he independently solved that problem and successfully applied it to computerized composite imaging. When British physicist Godfrey Newbold Hounsfield led the team of radiologists that built the first clinical CT scanner in 1971, he was unaware of either Radon or Cormack. Many improvements of CT have appeared since the 1970s, including positron emission tomography (PET) and single photon emission computed tomography (SPECT).

Magnetic resonance imaging

In the 1990s magnetic resonance imaging (MRI) and PET enabled ultrasophisticated, computer-driven brain mapping projects, such as that directed by Arthur Toga at the University of California at Los Angeles. MRI is safer than PET because it does not expose patients to radiation. MRI was originally called nuclear magnetic resonance (NMR) because it resonates hydrogen nuclei in the body, but in 1983 the American College of Radiology voted to change the name to avoid misunderstandings within the burgeoning "No Nukes" movement. A refinement of MRI, magnetoencephalography (MEG), measures brain activity directly rather than inferring it from relative oxygen levels, electrical impulses, and other intracranial data.

See also **Angiography; Electroencephalogram (EEG); Nuclear Magnetic Resonance (NMR, MRI); Positron Emission Tomography (PET); Psychiatry, Diagnosis and Non-Drug Treatments; Tomography in Medicine; X-Rays in Diagnostic Medicine**

ERIC V.D. LUFT

Further Reading

Bucy, P.C., Ed. *Modern Neurosurgical Giants*. Elsevier, New York, 1986.
Clarke, E. and Dewhurst, K. *An Illustrated History of Brain Function: Imaging the Brain from Antiquity to the Present*. Norman, San Francisco, 1996.
Clarke, E. and O'Malley, C.D. *The Human Brain and Spinal Cord: A Historical Study*. University of California Press, Berkeley, 1968.
Clifford, R.F. and Bynum, W.F., Eds. *Historical Aspects of the Neurosciences: A Festschrift for Macdonald Critchley*. Raven, New York, 1982.
Corsi, P., Ed. *The Enchanted Loom: Chapters in the History of Neuroscience*. Oxford University Press, New York, 1991.
Ecker, A.D. *The Normal Cerebral Angiogram*. Charles C. Thomas, Springfield, 1951.
Fox, W.L. *Dandy of Johns Hopkins*. Williams & Wilkins, Baltimore, 1984.
Krayenbühl, H.A. and Yasargil, M.G. *Cerebral Angiography*. J.B. Lippincott, Philadelphia, 1968.
Taveras, J.M. Diamond Jubilee Lecture: Neuroradiology, Past, Present, and Future. *Radiology*, 175, 593–602, 1990.

Nitrogen Fixation

In 1898, the British scientist William Crookes in his presidential address to the British Association for the Advancement of Science warned of an impending fertilizer crisis. The answer lay in the fixation of atmospheric nitrogen. Around 1900, industrial fixation with calcium carbide to produce cyanamide, the process of the German chemists Nikodemus Caro and Adolf Frank, was introduced. This process relied on inexpensive hydroelectricity, which is why the American Cyanamid Company was set up at Ontario, Canada, in 1907 to exploit the power of Niagara Falls. Electrochemical fixing of nitrogen as its monoxide

was first realized in Norway, with the electric arc process of Kristian Birkeland and Samuel Eyde in 1903. The nitrogen monoxide formed nitrogen dioxide, which reacted with water to give nitric acid, which was then converted into the fertilizer calcium nitrate. The yield was low, and as with the Caro–Frank process, the method could be worked commercially only because of the availability of hydroelectricity.

In Germany, BASF of Ludwigshafen was interested in diversification into nitrogen fixation. From 1908, the company funded research into nitrogen fixation by Fritz Haber at the Karlsruhe Technische Hochschule. Haber specialized in the physical chemistry of gas reactions and drew on earlier studies started in 1903 on the catalytic formation of ammonia from its elements, nitrogen and hydrogen. He attacked the problem with high pressures, catalysts, and elevated temperatures. Even under optimum conditions the yield was low, around 5 percent, but Haber arranged for unreacted hydrogen and nitrogen to be recirculated. Though exothermic, the reaction was carried out at 600°C in order to increase the rate. The preferred catalyst was either osmium or uranium. The main part of the apparatus was the furnace (later known as a converter) in which the gases were preheated by the outgoing reaction mixture. At a pressure of 200 atmospheres the gases were forced to react in the presence of the catalyst. Cooling moved the equilibrium in the direction of producing ammonia, which was liquified and separated from unreacted hydrogen and nitrogen.

On July 2, 1909, BASF catalyst expert Alwin Mittasch was convinced of the potential when he observed the benchtop reactor at work. Patents were filed in Germany and elsewhere, and Haber came to an agreement with BASF over royalties. BASF chemist Carl Bosch confronted the difficulties of scaling up Haber's 0.75-meter-high converter to a pilot plant. BASF had to seek cheaper catalysts than osmium and uranium, but with similar levels of activity, to build reactors to withstand high temperatures and pressures, and to establish inexpensive sources of nitrogen and hydrogen.

Mittasch undertook catalyst experiments in miniature high-pressure tubes. Suitable catalysts based on iron compounds, such as a Swedish magnetite, were available just by chance. In 1910, an iron–aluminum catalyst with activity close to that of osmium and uranium was chosen. A stable iron catalyst with aluminum and potassium used as the promoter was found to be successful in 1911, and Mittasch soon added calcium as a third promoter. In this way, he came up with the catalyst that would be favored in industrial synthesis of ammonia.

The main step forward was achieved in 1911, with Bosch's double-wall converter. The inner wall was made of low-carbon soft steel; the outer wall was of ordinary steel. Hydrogen diffusing through the inner wall underwent loss of pressure and then came into contact with the outer wall, which though hot from external heating did not become brittle. Stress on the outer wall was reduced considerably, particularly when a little hydrogen was allowed to escape into the atmosphere through small holes, called Bosch holes. Internal heating of the converter was by combustion of gas, later replaced by internal electrical heating.

In 1910, hydrogen became available from the reaction between steam and red-hot coke, which produced hydrogen and carbon monoxide (water gas). At first, nitrogen was available from the Linde process for liquefaction of air. The most important part of the physical processing of the gases was their compression. To achieve this under previously untried conditions, Bosch had to seek out powerful leak-proof gas compressors. The recycling of unreacted gases introduced novel approaches to optimization and control loops. Monitoring of the various physical processes and the chemical reaction that took place in the reactor required instruments for measuring gas flow at high pressures, gas density, and product mixture composition. Fast-acting magnetic shut-off valves were also required.

The BASF ammonia factory, at Oppau, near Ludwigshafen, was opened in September 1913. The converter was 8 meters high, and weighed 8.5 tons. A much cheaper source of nitrogen was available from the action of air on hot coke that gave producer gas. Later, a mixture of nitrogen and hydrogen in the proportions required by the chemical equation (1:3) was made from producer gas, water gas and steam. By 1915, ammonia converters of 12 meters in height and 75 tons in weight were in operation. In the spring of that year, synthetic ammonia was converted into nitric acid for munitions production in Germany. The process was soon named Haber–Bosch (see Figure 4).

Luigi Casale and Giacomo Fauser in Italy and Georges Claude in France independently invented high-pressure ammonia processes. The Italian and French processes differed from the Haber–Bosch process in that they employed higher pressures and different catalysts. Casale's process of 1924 operated at 650 to 750 atmospheres and incorporated a circulating system similar to that used in the

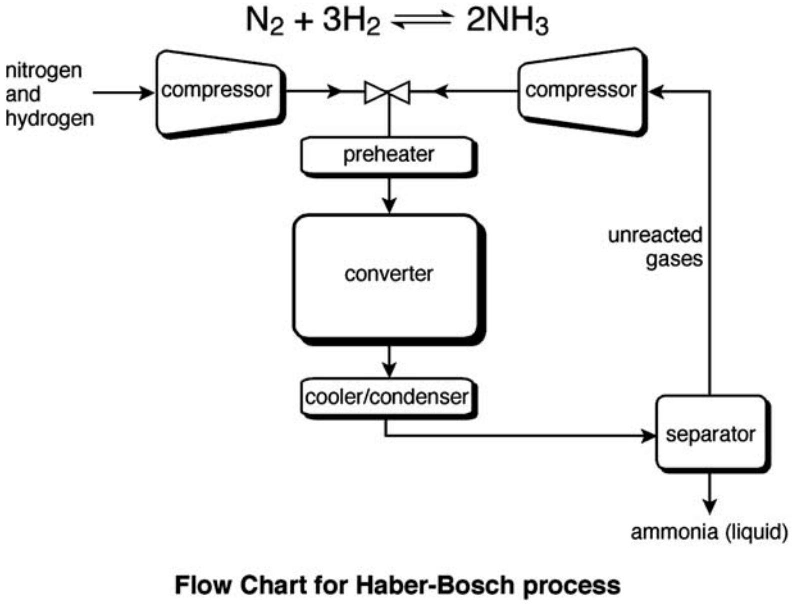


Figure 4. Flow chart showing the Haber–Bosch process for producing ammonia.

Haber–Bosch process. Fauser, a consulting engineer, used a 250-millimeter cannon as reactor in the garden of his home. From 1925, his studies were backed by Montecatini, and a viable ammonia process operating at 300 atmospheres was established. Georges Claude investigated high-pressure ammonia synthesis from 1917 using pressures of around 1000 atmospheres. Because of the high pressure, only 10 percent of unreacted nitrogen and hydrogen remained after passage through four convertors in series.

An estimate of world ammonia capacity for 1932–1933 showed that the Haber–Bosch process represented 53 percent of the total. Haber's reward was directorship of the new Kaiser Wilhelm-Institut für Physikalische Chemie und Elektrochemie in Berlin in 1912 and the Nobel Prize in 1918. Bosch's role in high-pressure synthesis was acknowledged with the Nobel Prize, received jointly with Friedrich Bergius, in 1931.

See also **Fertilizers**

ANTHONY S. TRAVIS

Further Reading

Bosch, C. *Nobel Lectures: Chemistry, 1922–1941*. Elsevier, Amsterdam, 1966, pp. 326–340.

Haber, L.F. *The Chemical Industry 1900-1930: International Growth and Technological Change*. Clarendon, Oxford, 1971.

Reader, W.J. *Imperial Chemical Industries: A History*, vol. 1: *The Forerunners, 1870–1926*. Oxford University Press, London, 1970–75. Includes difficulties in imitating the Haber–Bosch process at ICI, and how they were overcome.

Reinhardt, C. Über Wissenschaft und Wirtschaft. Fritz Habers Zusammenarbeit mit der BASF 1908 bis 1911, in Naturwissenschaft und Technik in der Geschichte, 25 Jahre Lehrstuhl für der Naturwissenschaft und Technik am Historischen Institut der Univesität Stuttgart, Albrecht, H., Ed. Verlag für Geschichte der Naturwissenchaften und der Technik, Stuttgart, 1993, pp. 287–315. A source of references to the German literature on Haber's ammonia work, including historical and historiographical accounts.

Smil, V. *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production*. MIT Press, Cambridge, MA, 2001.

Stoltzenberg, D. *Fritz Haber: Chemiker, Nobelpreisträger, Deutscher, Jude* VCH, Weinheim, 1994. Nitrogen fixation is covered in chapter 5, "Die Fixierung des Stickstoffs," pp. 133–197.

Haber, L. F. *The Poisonous Cloud: Chemical Warfare in the First World War*, Clarendon, Oxford, 1986. Studies on Haber normally place great emphasis on his controversial gas warfare work.

Travis, A.S. High pressure industrial chemistry: the first steps, 1909–1913, and the impact, in *Determinants in the Development of the European Chemical Industry, 1900–1939: New Technologies, Political Frameworks, Markets and Companies*, Travis, A.S., Homburg, E., Schröter, H. and Morris, P.J.T., Eds. Kluwer, Dordrecht, 1998, pp. 1–21.

Nuclear Fuels

Nuclear fuels are the propellants of nuclear reactors. In the context of the nuclear industry, they are divided into fertile materials and fissile materials. The former are treated radioactive minerals, while the latter are produced in nuclear reactors or through chemical separation with the fertile materials. In nature there are two types of

fertile materials, natural uranium (uranium-238) and thorium (thorium-232). The first is used to produce the fissile isotopes uranium-235 and plutonium-239. The second is used to obtain the fissile uranium isotope uranium-233. It is also possible to reprocess plutonium to obtain another type of fissile plutonium isotope, plutonium-241.

In 1789, the German Martin H. Klaproth, professor of chemistry at the University of Berlin, isolated natural uranium in the form of uranium dioxide (UO_2). Klaproth succeeded in extracting the mineral from a sample of pitchblende in the Joachimsthal ores (Czechoslovakia). This was named uranium in honor of William Herschel's discovery of the planet Uranus eight years before. While Henri Bequerel's experiments with uranium salts in 1896 demonstrated that uranium was naturally radioactive, the fissile properties of uranium were not recognized until 1938. The German physical chemists Otto Hahn and Fritz Strassmann demonstrated that neutron-bombarded uranium produced minimal quantities of lighter elements. Hence, uranium nuclei could undergo fission with the liberation of neutrons, radiation, and energy. In certain conditions, fission-produced neutrons could fission other uranium nuclei and thereby activate a self-sustaining reaction. But in February 1939, a joint paper by the Danish physicist Niels Bohr and his American colleague John Wheeler clarified that only a small fraction (1/139) of natural uranium (i.e., uranium-235), was responsible for the fission. The uranium-235 nucleus contains an even number of protons and an odd number of neutrons (even–odd nucleus). This asymmetry makes its binding energy smaller than the uranium-238 isotope (odd–odd nucleus). Thus Bohr and Wheeler concluded that uranium-238 captures neutrons, while uranium-235 undergoes fission.

From 1940 onward, several processes of separation were conceived in order to isolate the fissile uranium-235 from natural uranium and produce it at industrial scale. In 1940, the physical chemist Francis Simon designed the gaseous diffusion process at the University of Oxford in the U.K. Simon proposed to transform uranium oxide into a gaseous form, uranium hexafluoride (UF_6) and then pass it through a metallic membrane punctured with millions of microscopic holes. The lighter $^{235}UF_6$ would pass through the membranes with higher speed and so be isolated. The American physicists John Dunning and Eugene Booth of Columbia University adopted the process in 1941 and designed the first plant for separation based on the gaseous diffusion process. Construction of the plant began in 1943 at the Clinton Engineering Works in Tennessee in the U.S., and by the summer of 1945, it was successfully in operation.

Another process of enrichment was electromagnetic isotope separation, first analyzed by the American physicist Ernest O. Lawrence at the Radiation Laboratory of Berkeley. In the electromagnetic process, ions of gaseous uranium compound move in a circular ring due to the action of a strong magnetic field. Then the ions separate into two beams, the lighter uranium-235 ions following a narrower arc than the uranium-238 ions. In 1942, Lawrence designed the CALUTRON (California University cyclotron) that was eventually built in many models in the U.S. after 1943.

Finally, a third process of enrichment was the thermal separation of uranium, analyzed in detail by the American physicist Philip Abelson, who designed the first plant for thermal separation in 1943. In thermal separation the volatile uranium hexafluoride compound $^{235}UF_6$ is separated from the heavier UF_6 by heating.

Although these industrial methods for production of enriched uranium were aimed at the production of fissile material for atomic weapons, their concept was adopted by the nuclear industry in the 1950s when uranium became a fuel for commercial generation of electricity. The gaseous diffusion process proved by far the most successful, and 98 percent of enriched uranium is currently produced this way.

Plutonium does not exist in nature. It was artificially produced for the first time in January 1940 by bombarding uranyl nitrate hexahydrate (UNH_6) with an intense emitter of α-rays in a 9-inch (230-millimeter) cyclotron. The American chemist Glenn T. Seaborg at the Radiation Laboratory at Berkeley named the unknown chemical element plutonium after the planet Pluto, discovered in 1930.

The potentialities of plutonium (Pu) as a fissile material were thoroughly investigated between 1940 and 1943 and believed to be even higher than those of enriched uranium. Seaborg noted that plutonium could be obtained from natural uranium, which captures bombarded neutrons (^{239}U), decays into neptunium-239 (^{239}Np, 23.5 minutes), and then to plutonium-239 (2.3 days). In 1943, Seaborg also developed the chemical method of separating plutonium from irradiated uranium in which the uranium would be put in a chemical carrier and then undergo several processes of oxidation and reduction. The Clinton laboratories were the first research facility in which a new pile and a new chemical separation plant were built for

the experimental production of plutonium from natural uranium. In 1944 a new complex of reactors and chemical extraction plants were erected in Hanford, Washington, by Pasco for the first large-scale production of plutonium. Similarly to uranium, plutonium was initially produced for military purposes. From the 1950s onward, plutonium was used also as fissile material in nuclear reactors for electricity production.

Thorium was first discovered and extracted as a metal by the Swedish chemist Jöns J. Berzelius (1828), who named it after Thor, the Nordic God of thunder. The main source of thorium is the mineral monazite. Thorium's fissile properties were dismissed by the 1939 Bohr–Wheeler paper, but in December 1940 the Swiss physical chemist Egon Bretscher predicted that thorium would produce the fissile isotope uranium-233. In 1942, Seaborg observed that thorium-233, similarly to natural uranium, captures neutrons (^{234}Th) and decays to protactinium-233 (^{233}Pa, 22.2 minutes) to uranium-233 (27.4 days). Uranium-233 is a less valuable fissile material compared touranium-235 and plutonium-239. This ruled out its use as nuclear bomb material in World War II, but the abundance of natural thorium compared to natural uranium fostered its use in nuclear reactors for peaceful purposes beginning in the 1950s.

See also **Fission and Fusion Bombs; Particle Accelerators: Cyclotrons, Synchrotrons, and Colliders; Nuclear Reactors**

SIMONE TURCHETTI

Further Reading

Finch, W.I. *Uranium, Its Impact on the National and Global Energy Mix and Its History, Distribution, Production, Nuclear Fuel-Cycle, Future, Relation to the Environment*. U.S. Geological Survey, Washington, 1997.

Gowing, M. *Britain and Atomic Energy, 1939–1945*. Macmillan, London, 1964.

Kathren, R.L., Gough J.B. and Benefiel G.T., Eds. *The Plutonium Story, The Journals of Professor Glenn T. Seaborg, 1939–1946*. Battelle Press, Columbus, 1994.

Rhodes, R. *The Making of the Atomic Bomb*. Penguin, New York, 1986.

Smith, E., Fox, A.H., Sawyer, T. and Austin, H.R. *Applied Atomic Power*. Blackie, London, 1946.

Ursu, I. *Physics and Technology of Nuclear Materials*. Pergamon Press, Oxford, 1985.

Nuclear Magnetic Resonance (NMR) and Magnetic Resonance Imaging (MRI)

The phenomenon of nuclear magnetic resonance (NMR) was discovered independently in 1946 by Felix Bloch at Stanford University, and by Edward Purcell at Harvard University. For this they were jointly awarded the 1952 Nobel Prize in Physics. Chemists were quick to see the potential of NMR, as the NMR signal gives valuable information about the structure of molecules. This is because the atomic electrons, which determine the chemical properties of materials, interact with the nuclei which give rise to the NMR signals. Much later, in 1973, Paul C. Lauterbur outlined how NMR could be used to form images for medical diagnosis. Raymond Damadian's 1971 finding that cancerous tissue could have different NMR properties from normal tissue had prompted his work. This eventually led to a major new application of NMR in medical imaging, which we now call magnetic resonance imaging (MRI). Although Damadian and Lauterbur were both working in the U.S., much of the pioneering work needed to turn the revolutionary idea into a practical reality was done in the U.K. in the 1970s and 1980s, at the Universities of Nottingham and Aberdeen and at the EMI Company's London research laboratories. In 2003, the Nobel Prize in Medicine was awarded to Lauterbur, University of Illinois at Urbana, and Sir Peter Mansfield, University of Nottingham, for their discoveries, emphasizing the diagnostic importance and widespread use of NMR.

Atomic nuclei are positively charged, and some (but not all) have the quantum mechanical property termed spin. If an object is both charged and spinning, it will generate a magnetic field in the same way that a current circulating round a loop will generate a magnetic field. Thus, a nucleus that has spin can be thought of as being a tiny bar magnet. Normally, nuclear spins do not have any preferred direction of alignment. However, if they are placed in a strong magnetic field they will tend to align with it, in much the same way as a set of compass needles align with the Earth's magnetic field. The alignment brought about by the strong magnetic field produces an observable nuclear magnetism. Until the nuclei are placed in the strong magnetic field their magnetic dipoles are randomly oriented and their average effect is zero, so that the phenomenon of NMR cannot occur. In fact, even in a very strong magnetic field their alignment is only weak, as the nuclear magnetic moments are randomly disturbed by thermal agitation as if they were compass needles being violently shaken.

The gyroscope, or spinning top, is a good analogy for the behavior of the nuclear magnetization in a strong magnetic field. If the spinning top is perturbed from its initial alignment with the Earth's gravitational field while supported by a

table top, then it precesses around a vertical axis through its point of contact with the table. The precession is a much slower motion (in terms of revolutions per second) than the spin of the top around its own axis. Similarly, if the nuclear magnetic dipoles are perturbed from their alignment with the magnetic field, they precess around it, as shown in Figure 5. The precession frequency f_0 is often referred to as the Larmor frequency (after Joseph Larmor, an Irish physicist who investigated the behavior of electrons in magnetic fields at the end of the nineteenth century) and is given by the equation $f_0 = \gamma B_0$. Different nuclei have different values of γ. Protons have a gyromagnetic ratio of 42.6 megahertz per tesla, while the figure for sodium is 11.3 megahertz per tesla. Thus protons (hydrogen nuclei) precess at 42.6 megahertz in a magnetic field of strength 1 tesla, while sodium nuclei (specifically of the sodium isotope with atomic weight 23) precess at only 11.3 megahertz. It is possible to determine the frequency of a signal very accurately, by comparing it to a stable reference signal of known frequency. Thus NMR can be used to make a highly precise magnetometer—a device for measuring magnetic field strengths very accurately.

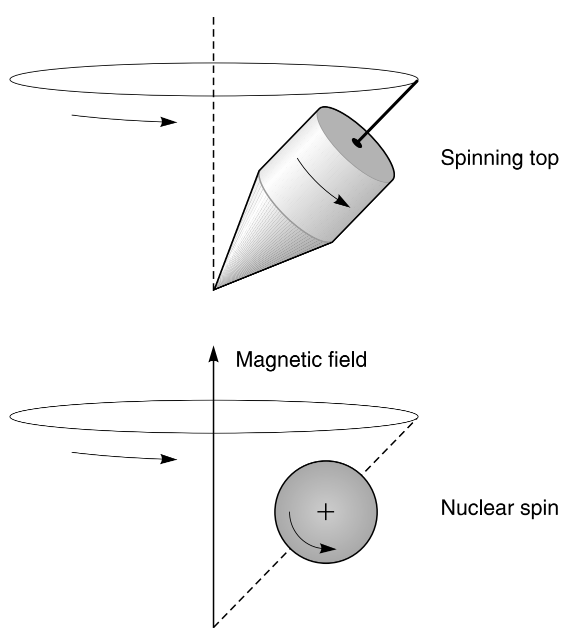


Figure 5. The spinning top is a good analogy to the behavior of the nuclear magnetism in a strong magnetic field. Once the top is pushed away from the vertical, it precesses around the gravitational field of the Earth. Similarly, once the spinning nuclei are pushed away from their alignment with the magnetic field, they also precess. The rate of precession is proportional to the strength of the magnetic field.

Whereas the spinning top can be pushed from the vertical by a tap of the finger, the nuclear spins have to be pushed by an oscillating magnetic field applied at right angles to the main magnetic field. The oscillation frequency has to equal the Larmor precession frequency f_0, or nothing happens. This is what is meant by resonance. The field strength used in a typical NMR or MRI machine is a few tesla, so that the Larmor frequency is in the radio-frequency (RF) range. The RF magnetic field is applied by means of a tuned RF coil surrounding the patient's body or head, with power supplied by a radio-frequency power amplifier.

The first NMR experiments in the late 1940s used the "continuous wave" technique, in which the frequency of the RF field is steadily increased (or decreased) while passing through resonance. In 1950 Erwin Hahn proposed the "pulsed" method of NMR in which the entire frequency response is obtained following a short powerful burst of transmitted RF energy called an RF magnetic field pulse. The difference in methods can be understood by considering two possible methods of testing a church bell. In the analogy for the continuous wave method, a loudspeaker is used to produce a pure note, the frequency of which is steadily increased. When the natural tone of the bell is reached, the bell will begin to vibrate in sympathy with the applied sound. The pulsed method is faster and more direct, and can be likened to striking the bell, and listening to its note as the sound dies away. The note contains a mixture of all the natural frequencies of the bell. Once the nuclei are precessing, their magnetic fields are also precessing, so that a tiny voltage is induced in the RF coil by the principle of electromagnetic induction. The frequency of this oscillating voltage equals the precession frequency. If a number of different frequencies are present in the signal, because the nuclei are in a number of different electronic environments, then the different frequency components have to be separated by a mathematical process called Fourier transformation, carried out by a computer.

The use of NMR to investigate the atomic and molecular structure of materials is called NMR spectroscopy. During the 1950s and 1960s, NMR spectroscopy became a widely used technique for the nondestructive analysis of small samples, particularly of liquids. The electron cloud surrounding the nuclei within individual molecules modifies the strength of the magnetic fields sensed by the NMR sensitive nuclei, and hence changes the frequency of the NMR signal that the nuclei emit. In addition, neighboring nuclei can also

influence each other. Pulses of RF energy are used to perturb the NMR sensitive nuclei, and sensitive RF receivers are used to pick up the signals they give out. The pulsed method allied with computerized Fourier transformation revolutionized NMR spectroscopy in the 1970s. The "spectrum" is a plot of signal strength versus NMR frequency and contains much useful information about the chemical structure of the material under testing. NMR spectroscopy is now widely used in the fields of biomaterials, polymer chemistry, and solid-state physics.

Pulsed NMR techniques can also be used to form medical images. MRI normally uses signals arising from hydrogen nuclei, because they are so much more abundant in the body than any other NMR-sensitive nucleus and therefore give a measurable signal even from small volumes of tissue. However, phosphorus and sodium MRI is also possible. Hydrogen nuclei that are MRI visible occur predominantly in water in the tissues and in body fat. Figure 6 shows a patient lying on a table, about to be moved into an MRI scanner, with the head resting inside the RF head coil. The electromagnet has a large tunnel in which the patient lies and uses a special wire immersed in liquid helium so that the wire superconducts. Thus the magnet does not need any electrical power to generate the field. To form an image, it is essential to have a method of determining the position of the nuclear spins within the magnet. This is accomplished by field gradients, a method proposed by Paul Lauterbur of the State University of New York at Stony Brook in 1973. A field gradient coil modifies the strength of the main magnetic field along a particular direction so that it varies in a linear way. Three independent field gradient coils are used in order to generate the gradients in x, y, or z directions (along the magnet tunnel, left to right across the tunnel, and vertically). When a field gradient is on, therefore, the Larmor frequency will also vary in a linear fashion along the direction of the field gradient.

The first step in imaging is to tip the nuclear spins away from their alignment along the main magnetic field using a short-pulsed RF magnetic field oscillating at the Larmor frequency, as explained above. If the pulse is applied in the presence of a field gradient along the patient (along z), then only one transverse plane across the body will respond, as only one plane has a Larmor frequency that matches the RF frequency. The imaging process is now simplified to imaging a two-dimensional slice in an x–y plane, rather than imaging the entire three-dimensional body. The gradient along the patient is switched off, and a gradient switched on across the body (along x). This makes the nuclei precess faster on one side of the body, and more slowly on the other. Thus the frequency of the NMR signals varies in a linear way across the body, so that a particular NMR signal frequency corresponds to a particular left to

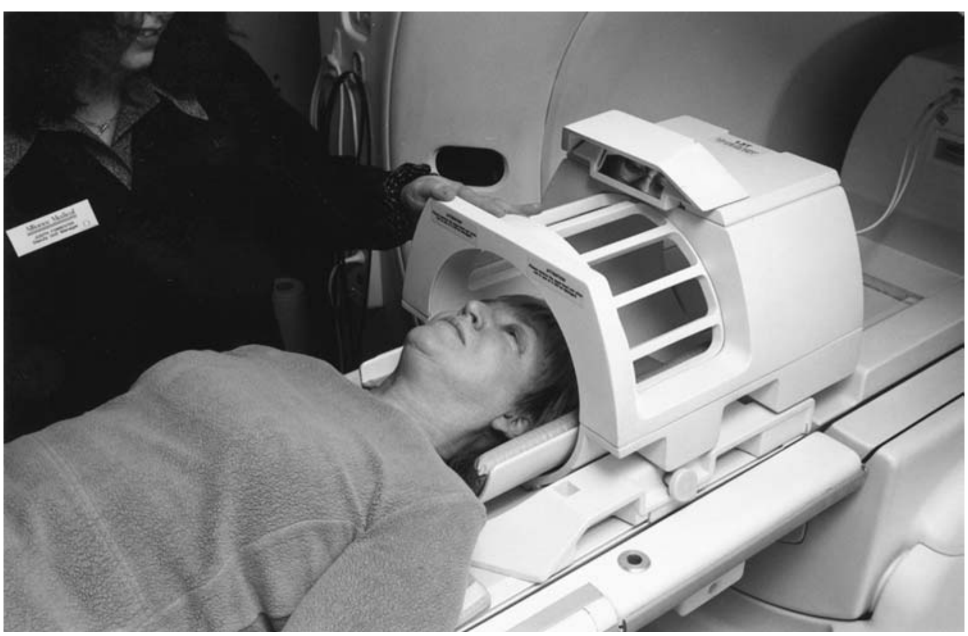


Figure 6. A patient lies on a table before being transferred into the MRI scanner. Her head is being imaged and has to be placed inside the radio-frequency head coil beforehand.

right (x) position. In order to form the image, the vertical (y) position also has to be encoded onto the NMR signal. Although a full discussion is impossible here, the technique involves repeating the above process many times, with a small pulse applied to the vertical y gradient just before the signal is collected and with the amplitude of the pulse being stepped up between each repetition. This is the "spin-warp" imaging method, invented by William Edelstein and James Hutchison of Aberdeen University in 1980. It is still the most widely used imaging method today. Modern MRI scanners rely on fast computers to carry out the large number of digital Fourier transforms necessary to form the images and to control very precisely the timing of the pulsed currents flowing in the gradient and RF coils.

Figure 7 shows a midline slice through the center of a human head on an MRI image. Note the excellent anatomical detail achieved. Areas with no protons, such as the air-filled sinuses behind the nose, give no signal and appear dark. Dense hard bone in the skull also has few protons and gives no signal. Watery fluid such as the cerebrospinal fluid around the spinal cord gives a different signal (brighter in this case) from that of the soft tissue of the brain because the NMR properties of the water protons are affected by their biophysical environment. In other words, image detail arises from tissue structure on a microscopic level, as well as from differences in proton density. This is the particular strength of MRI in medical diagnosis because many disease processes, including cancer and multiple sclerosis, and conditions such as infection and hemorrhage alter tissue structure and can therefore be visualized.

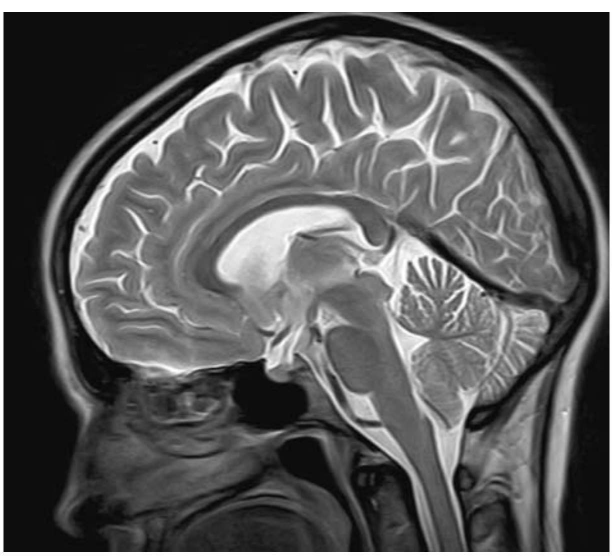


Figure 7. A section of the brain imaged by MRI. Note the excellent anatomical detail the scan gives.

A recent development in MRI is its use to repeatedly image the brain every 2 or 3 seconds as it performs different tasks or responds to various stimuli, a technique called functional MRI (fMRI), developed independently by Seiji Ogawa of AT&T Bell Laboratories in New Jersey and John Belliveau at Massachusetts General Hospital in Boston in the early 1990s. This requires a high-performance scanner capable of performing the very high-speed imaging technique of "echo-planar" imaging, invented by Peter Mansfield at Nottingham University in 1977. The fMRI technique relies on the oxygen-dependent magnetic effect of the iron atoms contained in the hemoglobin molecules of the blood. When part of the brain is active, its oxygen consumption increases, thus stimulating a large increase in the local blood supply, and the oxygen concentration in the tiny blood-filled capillary vessels is raised. For the particular rapid-imaging method used in fMRI, this causes a small increase in the signal detected. In its simplest form, fMRI alternates periods of rest, lasting 30 seconds or so, with equal periods of activation. Relatively large signal changes of a few percent can be obtained with visual activation using flashing lights or rapidly alternating checkerboard patterns (black squares turning to white and vice-versa), while more subtle cognitive tasks involving memory or reasoning give smaller changes. In other words, MRI can image your thoughts, an application never dreamt of by Felix Bloch and Edward Purcell in 1946.

See also **Cardiovascular Disease, Diagnostic Methods; Neurology**

Thomas W. Redpath

Further Reading

Callaghan, P.T. *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford University Press, Oxford, 1991.

Farrar, T.C. and Becker, E.D. *Pulse and Fourier Transform NMR*. Academic Press, Orlando, 1971.

Gadian, D. *NMR and its Applications to Living Systems*, 2nd edn. Oxford University Press, Oxford, 1995.

Hore P.J. *Nuclear Magnetic Resonance*. Oxford University Press, Oxford, 1995.

Redpath, T.W. MRI developments in perspective (review). *Br. J Radiol.*, 70, s70–80, 1997.

Rinck, P.A. *Magnetic Resonance in Medicine*. Blackwell Wissenschafts-Verlag, Berlin, 2001.

Stark, D.D. and Bradley, W.G. *Magnetic Resonance Imaging*, 3rd edn. Mosby, St. Louis, 1999.

Nuclear Reactor Materials

Over the latter half of the twentieth century, several materials were used by the nuclear industry in the construction and functioning of nuclear reactors. Those materials have been divided into the following categories: moderators, structural materials, coolants, and shielding materials.

The moderator is used in nuclear reactors to slow neutrons down to fission energy. Slowed down or ''thermal'' neutrons have a far higher probability to cause nuclear fission than fast neutrons; the use of a moderator is therefore essential in nuclear reactors that exploit thermal neutrons (thermal reactors). Moderation is obtained through repeated elastic collisions of the neutrons with the moderator's nuclei and the transfer of kinetic energy from the first to the second.

Despite initial attempts to use light water as a moderator in early experiments with nuclear piles, graphite appeared far more reliable and constituted the main type of moderator for a long time. The use of graphite is widespread. Its manufacturing process was first outlined in 1896 by the American industrialist Edward G. Acheson. He used petroleum coke and coal-tar pitch in a large furnace capable of reaching temperatures of 3000°C. When in 1942 the Italian physicist Enrico Fermi designed the first experimental nuclear reactor, CP-1, at the University of Chicago, he was advised by his Hungarian colleague Leo Szilard to try using graphite as a moderator. In December 1942, Fermi succeeded in reaching criticality with the graphite pile CP-I. This success influenced future designs of nuclear piles, although the low moderating power of graphite implied the design of big power plants with large cores and therefore deployment of large quantities of structural materials. Graphite is still in use today in nuclear reactors of the graphite core reactor (GCR) type.

In 1943, graphite's properties were thoroughly examined by the theoretical physicist Eugene Wigner at the University of Chicago. He argued that radioactivity heavily affects graphite and modifies its strength and thermal conductivity. Eventually several scientists and industrialists within the nuclear establishment considered the use of other types of moderator such as heavy water (D_2O) and light water (H_2O). Heavy water was produced for the first time in 1934 by the British Imperial Chemical Industries (ICI) through a process of water electrolysis. Previously, the American physical chemist Harold Urey had isolated the isotope deuterium (^2H or D). The first experimental heavy water pile CP-3 was built in 1944 at the U.S. Argonne National Laboratory. It consisted of a tank filled with 6.5 tons of heavy water containing 121 rods of uranium. Although the Argonne CP-3 was the first heavy water reactor, the technology associated with this type of moderator was mainly developed in Canada. From 1943 British, Canadian, and French scientists gathered in Montreal as members of the Anglo–Canadian project code-named ''Tube Alloys.'' Their purpose was to build a heterogeneous pile in which uranium rods were immersed in heavy water. In 1945 they built the prototype zero energy experimental pile (ZEEP) at the new Chalk River Nuclear Laboratories (Canada), and in 1947 the heavy-water nuclear reactor X (NRX) was brought into operation. Heavy water is still used in the Canadian-designed commercial reactors of the CANDU type. Heavy water reactors have a larger output compared with graphite reactors, and they have reduced dimensions. Their main drawback lies in the production of heavy water, which is still a very expensive and cumbersome process.

After being initially dismissed as a moderator, light water was reconsidered because its moderating power is higher than both graphite and heavy water. Light water was used in reactors of the light water reactor (LWR) type from the 1950s onward.

Structural materials are used to make the core and several other essential parts of a nuclear reactor. Thus, they must possess a high resistance to mechanical stress, stability to radiation and high temperature, and low neutron absorption. Materials used for structural purposes are aluminum, beryllium, carbon (graphite), chromium, iron, magnesium, nickel, vanadium and zirconium. Stainless steels, alloys of iron, chromium, nickel, and other metals, are by far the most used structural materials since the beginning of the nuclear industry.

In the nineteenth century, the experiments conducted by the British natural philosopher Michael Faraday with iron and nitric acid had shown that iron alloys had a very high corrosion resistance (or passive state). The characteristics of iron alloys were thoroughly examined at the beginning of the twentieth century. In 1911, P. Monnartz investigated and classified the Fe–Cr (ferritic) alloys. He also developed a number of methods (e.g., introduction of molybdenum) to enhance their passive state. In 1912, the Germans Eduard Maurer and Benno Strauss, chemists of the Krupp research laboratory, patented a method for the treatment of Fe–Cr–Ni alloys. From 1914 those alloys became the main structural materials for

chemical plants, such as for ammonia or nitric acid production. In the same year, the English metallurgist Henry Brearley used iron alloys to make "rustless" cutlery, hence the name "stainless" steels. In 1931, Brearley synthesized a new type of stainless steel—the austenitic—using chromium and nickel. From the end of World War II, several new stainless steels were designed, introducing new elements such as titanium, silicon, and nitrogen. Since then, stainless steels were applied as structural materials in the petrochemical industry, thus their use was suggested for the nuclear industry as well. Stainless steels are the main off-core structural materials for thermal reactors and are essential materials for fast neutron reactors. Their main drawback is that they absorb neutrons, which limits their in-core applications. Carbon and ferritic steels are used for nuclear pressure vessels in LWR and heavy water reactors (HWR). Lightly alloyed steels are used in PWR reactors. Ferritic steels were also used in the commercial reactors of the CANDU–PHRW type for the manufacture of pressure tubes and fittings.

Coolants are used in nuclear reactors to evacuate the thermal energy produced from the core. Materials used for cooling purposes are water (pressurized or boiling), heavy water, liquid metals (mainly sodium), or gases (carbon dioxide, helium, or nitrogen oxide). The process of energy transfer may imply a phase transition in the coolant (i.e., water storing heat through vaporization). The containment of radiation and heat within nuclear power plants is obtained through the shielding materials. Nuclear reactors are provided with an internal thermal screen made of water or heavy water and an external biological screen made of air, water, or concrete. In fast reactors, stainless steels are used for a thermal screen while sodium, iron, or boron is used as the biological screen.

See also **Nuclear Fuels; Nuclear Reactors, Thermal, Graphite-Moderated; Nuclear Reactors, Thermal, Water-Moderated; Nuclear Reactors, Fast Breeders**

SIMONE TURCHETTI

Further Reading

Dahl, P. F., *Heavy Water and the Wartime Race for Nuclear Energy*. Institute of Physics Publishing, Bristol, 1999.

Fleming, R.P.H. Design and development, structure and properties of some new stainless steels for high-temperature use, in *Mechanical Behaviour and Nuclear Applications of Stainless Steel at Elevated Temperatures: Proceeding of the International Conference, Varese, Italy, 1981*. Metals Society, London, 1982.

Gowing, M. *Britain and Atomic Energy, 1939–1945*. Macmillan, London, 1964.

Nightingale, R. E. *Nuclear Graphite*. Academic Press, New York, 1962.

Rhodes, R. *The Making of the Atomic Bomb*. Penguin, New York, 1986.

Streicher, M. A. Stainless steels: past, present and future, in *The Metallurgical Evolution of Stainless Steels*, Pickering, F.B., Ed. Metals Society, London, 1979.

Ursu, I. *Physics and Technology of Nuclear Materials*. Pergamon Press, Oxford, 1985.

Nuclear Reactors: Fast Breeders

The idea of a fast breeder reactor (FBR) was first conceived in 1946 by the Canadian physicist Walter H. Zinn at the Argonne National Laboratory in the U.S. On the basis of wartime developments in nuclear reactor research, Zinn thought a combination of two options within reactor technologies was feasible: fast neutron nuclear fission and the breeding principle. Fast reactors produce nuclear fission with fast neutrons rather than thermal neutrons. Fast neutrons prompt critical reactions with a large energy release in a short time and without a moderator operating in the core. Breeder reactors have a core of fissile material (i.e., uranium-235 or plutonium-239) produced in nuclear reactors or through chemical separation, and a blanket of fertile material (i.e., uranium-238), the treated radioactive mineral. Once in operation, breeder reactors incinerate the fissile material in the core and emit neutrons as fission products. Thus the blanket is neutron bombarded, the fertile material is irradiated, and afterward transformed into fissile material (i.e., plutonium-239) by neutron capture and following decay. In optimal operation conditions, the fissile material produced through breeding equals the fissile material incinerated in the core, so that the reactor perpetuates indefinitely the production of its fuel.

In 1948, Zinn designed a FBR provided with a good coolant, a liquid alloy of sodium and potassium (NaK), to promptly transfer the enormous heat produced in the core. This project soon received priority within the U.S. Atomic Energy Commission (AEC). The construction of EBR-I (experimental breeder reactor) started in June 1949 at the Idaho reactor research station and was completed in 1951. In December 1951 the reactor went critical, and EBR-I was the first nuclear reactor capable of producing electricity at the initial rate of 45 megawatts. In 1952, EBR-I proved to be a breeder reactor as part of the natural uranium in the blanket was irradiated and then transformed into plutonium. In the same year, a leak in the nuclear reactor heat exchanger caused

its temporary interruption. The leak highlighted the fact that the corrosive nature of sodium would be a major concern for FBR technology. In 1955 a combination of causes (operator error and a temperature effect) caused the meltdown of the EBR-I core. The core was eventually replaced, but the reactor was shut down in 1963. In that year, the Enrico Fermi I—a full-scale FBR cooled with liquid sodium—started operations in Lagoona Beach, Michigan, in the U.S. In 1966 during a normal test, the instrumentation registered an abnormal temperature in the core and an erratic rate of change in the neutron population. After the plant was shut down, it was discovered that two of the fuel assemblies had melted. Enrico Fermi I reached full power again only in 1970, and it was shut down afterwards. The two nuclear accidents, the fourth level of the International Atomic Energy Administration (IAEA) scale of nuclear events, represented a major drawback for the FBR American program.

The U.S. slowed down investments in FBR technology in 1970. However, FBRs were still considered very reliable in Europe, Japan, and the former Soviet Union, where they were regarded as the "nuclear system of the future." From 1969, the first French prototype Rapsodie (20 megawatts) was built by the Commissariat à l'Energie Atomique (CEA) at the Saclay research laboratory. In 1974, the FBR Phénix was built at Avignon (250 megawatts), and in 1984, the full-scale FBR Super Phénix was built (1200 megawatts) at Malville on the Rhone River. From the 1970s, the Soviet Union heavily invested in FBR technologies with two experimental FBRs (15 megawatts); an intermediate FBR (BN350, 150 megawatts, 1973); and a full-scale FBR (BN600, 600 megawatts). In 1977 the Dounreay PFR (270 megawatts) was the first FBR operating in Britain. In Germany in 1978 the prototype KNK Karlsrhue (18 megawatts) was operative, while a new FBR, the SNR 300 was built in 1987 (327 megawatts).

However, in thirty years the opinion of experts and the public on FBR technology changed as FBR did not prove to be efficient nuclear systems. The Soviet FBRs suffered serious sodium leaks and raised concern among nuclear reactor experts. The French Super Phénix operated at full power for just six months, and in 1997 its closure was announced as part of the new French government's plans. The British Dounreay PFR was shut down in 1994 after several small accidents occurred, and Britain generally opted for investing in FBR development at the European rather than British level. In Europe the development of a common program for a European Fast Reactor (EFR) capable of producing 1450 megawatts was evaluated but never fully accomplished.

The development of FBR technology was seriously evaluated after several accidents occurred in existing prototypes and commercial reactors. Those accidents have also slowed down the process of commercialization of FBR technology. The critics of these types of reactors point out that liquid sodium is a highly corrosive chemical compound that causes leaks, tube vibrations, and flow instabilities. Sodium also reacts with air and water and interacts with the fuel in emergency conditions. This is considered a major concern as the explosive nature of the interaction between fuel and coolant can lead to uncontrolled critical conditions in the rods. Moreover, technical problems that do not emerge at the level of prototypical FBRs affect full-scale commercial reactors because sodium instability increases when larger quantities of coolant are deployed. This implies corrections to original designs and escalating costs. For example, the French Super Phénix needed an expenditure of six times its original estimated cost to be completed.

Supporters of FBR technology argue that there is no reason to be concerned about liquid sodium's properties and behavior in nuclear reactors. There are no fundamental engineering problems connected with the construction and operation of FBRs. The main problem is simply the balance between incinerated fissile material and fissile material to replace it. In other words, FBRs incinerate fairly well but breed poorly, making the incineration/breeding balance very difficult to obtain. Among several positive aspects, FBR supporters stress that the breeding principle helps in confining radioactive waste. FBRs allow treatment of radioactive fissile material in the reactor establishment within 12 months rather than the transportation of radioactive material in and out from the reactor and its treatment in five years. Thus they provide a good way to confine nuclear fuel and radioactive materials. Furthermore, once the optimal incineration/breeding balance is obtained, the efficient incineration of plutonium is another favorable element of FBR technology as this helps to diminish the amount of plutonium, one of the most toxic chemical compounds in the world.

See also **Nuclear Reactors, Thermal Graphite-Moderated; Nuclear Reactors, Thermal Water-Moderated; Nuclear Fuels**

SIMONE TURCHETTI

Further Reading

Dombey, N. Fast reactors and problems in their development, in *The Fast Breeder Reactor. Need? Cost? Risk?* Sweet, C., Ed. Macmillan, London, 1980.
Hecht, G. *The Radiance of France, Nuclear Power and National Identity after WW2*. MIT Press, Cambridge, MA, 1998.
Hewitt, G.F. and Collier, J.G. *Introduction to Nuclear Power*. Taylor & Francis, New York, 2000.
Hewlett, R. and Duncan, F. *Atomic Shield: A History of the United States Atomic Energy Commission, vol. II, 1947–1952*. University of California Press, Berkeley, 1990.
Marshall, W. The UK fast breeder programme, in *The Fast Breeder Reactor. Need? Cost? Risk?* Sweet, C., Ed. Macmillan, London, 1980.
Mounfield, P. *World Nuclear Power*. Routledge, London, 1991.

Nuclear Reactors: Fusion, Early Designs

The production of nuclear energy through the fusion of two light chemical elements is better known as a controlled thermonuclear reaction (CTR). In the 1950s, explosive or uncontrolled thermonuclear reaction was achieved with the manufacture of hydrogen bombs, but CTR was never successfully accomplished. Among the several types of reactions considered, the following CTR has been attempted in fusion reactors:

$$^{2}\mathrm{H} + {}^{3}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + \mathrm{n} + 5{,}2 \times 10^{-13}\ \mathrm{J}$$

In the above reaction, deuterium (^{2}H or D) and tritium (^{3}H or T) fuse to release helium (He), neutrons, and energy (calculated in joules, J). Deuterium is contained in a small percentage in water, while tritium can be artificially produced from lithium by neutron bombardment in a breeding reactor.

In order to reach the fusion point, a gaseous mixture containing deuterium and tritium should be heated to 100,000,000°C and hold that temperature for enough time to activate a self-sustaining reaction. At elevated temperatures, a gaseous mixture becomes plasma, a state in which electrons and ions are no longer physically bonded. (The term plasma was first used in 1922 by the American physical chemist Irving Langmuir because the properties of a super-heated gas reminded him of blood plasma.)

Heating and confinement of plasma are the two main features of any fusion reactor. Plasma must avoid any contact with the walls of the vessel containing it in order to avoid the loss of temperature and subsequent instability that makes a controlled thermonuclear reaction impossible to achieve. Early designs of fusion reactors focused on confinement of plasma using magnetic fields.

In 1951, the American astrophysicist Lyman Spitzer at Princeton University designed the Stellarator, so called because thermonuclear reactions occur in stars and the device was aimed at reproducing the energy released when such reactions occur. The Stellarator was designed in three models—A, B, and C—between 1953 and 1961. Spitzer argued that "To keep the ions from hitting the wall, some type of force is required that will act at a distance. A magnetic field seems to offer the only promise." He designed a closed tube in the shape of a circular ring (torus) surrounded by external magnetic coils. In the device, the plasma was heated by electric current or radio waves and then introduced into the torus in vacuum conditions. Ideally, the coils should hold the plasma and move it in the ring for enough time to generate a CTR. The 1961 Stellarator C was the first prototype of a fusion reactor. This device reached temperatures of thousands of degrees Celsius and held the plasma for a few milliseconds, but it did not achieve CTR.

After 1952, another type of design was considered for fusion reactors: the pinch effect machine. Exploiting an idea originally developed in 1946 by his British colleagues George P. Thomson and Moses Blackman, the British physicist James Tuck devised the Perhapstron (to ridicule the Stellarator's "grandeur") at the U.S. Los Alamos National Laboratory. Perhapstron was a pinch effect machine in which an electric current is discharged in the plasma. In this way the current shapes a circular magnetic field that holds the plasma and pinches it into a thin filament. The problem of Tuck's device was the plasma's instability; that is, its capacity to hold the plasma by magnetic means. In 1955 Edward Teller addressed plasma instability as a major concern for CTR research. For this reason in 1957 a new Los Alamos pinch machine, Scylla, was designed. This was an "azimuthal" machine in which longitudinal wires around the pinch tube excited longitudinal currents in the plasma.

In August 1957 another experimental pinch prototype, the Zero Energy Thermonuclear Assembly (ZETA) started operation at the Harwell Atomic Research Establishment in Britain. This was the largest stabilized pinch machine ever built, and it was initially believed that CTR was finally achieved as neutrons were detected pouring out of ZETA. However, more accurate studies eventually clarified that the detected neutrons were not fusion products.

In the same year, Edward Gartner, leader of the Fusion Group at the U.S. Oak Ridge National Laboratory, established that a beam of ions trapped in a carbon arc and eventually sent into a pinch machine would enhance plasma stability. On this premise, two fusion reactor prototypes named Direct Current X (DCX) were successfully developed in the early 1960s. On the whole, pinch machines provided useful experimental data on CTR, but they never achieved fusion.

If pinch machines and stellarators focused on toroidal structures, the magnetic mirror machines were designed to attempt plasma confinement in straight tubes. In 1952, the American physicist Herbert York, head of the project for the new U.S. Lawrence Livermore National Laboratory, directed a research program for building a fusion device based on a linear structure. A young American PhD graduate, Richard F. Post, designed TableTop, a device in which the coils surrounding the straight tube would produce different magnetic fields. In that way the plasma would travel from the center to the ends of the tube and then back again, creating a mirroring, resonance effect. In 1954, the American Frederic H. Coengsen designed ToyTop, a machine to test plasma heating by magnetic compression. In 1960, ToyTop and TableTop were combined in the design of the large-scale mirror machine, ALICE (adiabatic low-energy injection and confinement experiment). The major drawback for mirror machines was "flute instability," the tendency of the plasma to avoid confinement in the straight tube. In order to avoid flute instability, the Russian physicist Abraham Ioffe proposed introducing magnetic wells at the ends of the tube. In 1957 another Russian physicist, Len Andreevich Artsimovich, designed OGRA, the first Soviet mirror machine manufactured at the Institute of Physics of Moscow and fitted with Ioffe's magnetic wells. Meanwhile, the British physicist Stephen Pease at the Culham Laboratory designed magnetic wells shaped as a tennis ball seam. The wells were eventually attached to the second Los Alamos ALICE prototype built in 1966.

During the 1960s, many fusion reactor projects went under critical review. This was largely because of the lack of success in harnessing fusion energy. Several projects were therefore withdrawn, such as the Livermore mirror machine Astron, designed by the Greek-born American physicist Nicholas C. Christofilos. In 1968, Artsimovich presented a new type of machine—the tokamak—whose innovative design would restore hope (and funding) in fusion reactors research.

See also **Fission and Fusion Bombs; Nuclear Reactors: Fast Breeders; Nuclear Reactors: Fusion, Later Designs (Tokamak)**

Simone Turchetti

Further Reading

Bromberg, J.L. *Fusion. Science, Politics and the Invention of a New Energy Source*. MIT Press, Cambridge, MA, 1982.

Hendry, J. and Lawson, J. *Fusion Research in the UK, 1945–1960*. AEA Technology, Harwell, 1993.

Herman, R. Fusion, *The Search for Endless Energy*. Cambridge University Press, Cambridge, 1990.

Hewitt, G. and Collier, J. *Introduction to Nuclear Power*. Taylor & Francis, New York, 2000.

Mounfield, P.R. *World Nuclear Power*. Routledge, London, 1991.

Stacey, W.M. Fusion: *An Introduction to the Physics and Technology of Magnetic Confinement Fusion*. Wiley, New York, 1981.

Nuclear Reactors: Fusion, Later Designs (Tokamak)

The tokamak reactor design aims to replicate in a terrestrial environment the conditions for fusion that exist in stars. There, the forces of high temperature and compression that prevent electrons from adhering to hydrogen nuclei and fusing the positively charged atoms (ions) into helium ions (a process that releases massive energy), are balanced by immense gravitational forces. In simulating these conditions, scientists use magnetic fields to stabilize, confine, and suspend ionized fuel (plasma) in a reactor vessel. This prevents the plasma from contacting the vessel walls, which would cool, and thus terminate, the reaction.

In the 1940s and 1950s, scientists developed two methods of magnetic plasma confinement. A magnetic field could be created within the plasma itself by running a current through it, producing a "pinch" effect. While initial "pinch" devices were tube-like, it was discovered that the plasma column cooled when it touched the ends of the cylinder. This led researchers to adopt the toroidal (hollow doughnut) reactor, within which the plasma formed a floating ring. The American physicist Lyman Spitzer fitted an electromagnet to such a reactor (the "Stellerator"), producing a toroidal magnetic field.

In the early 1950s, the Soviet physicists Andrei Sakharov and Igor Tamm proposed a reactor that generated both internal plasma and external toroidal magnetic fields. This concept was adopted by their colleague Lev Artsimovich in his T-3 reactor, the first "tokamak" (the Russian acronym for

toroidal chamber and magnetic coil), unveiled in 1968. The tokamak magnetic field is thus the combination of two magnetic fields: the stronger horizontal, toroidal field interacts with the weaker vertical, poloidal plasma field to produce a helical magnetic field. In confining its plasma for 0.01 to 0.02 seconds and heating it to 10,000,000°C, the T-3 produced results that suggested fusion energy was feasible.

The tokamak reactor subsequently became the standard tool for fusion research. The energy crisis of the 1970s resulted in state support for major projects in a number of industrialized countries including France, Japan, the U.K., and the U.S. The largest and most notable were the American Tokamak Fusion Test Reactor (TFTR), approved by the Atomic Energy Commission in 1974 and completed in 1982 at Princeton University, and the British–European Joint European Torus (JET), which began operations in 1983 in Culham, Oxfordshire, U.K. Other important tokamaks include Japan's JT-60 and General Atomics' DIII-D.

Plasma heating in fusion reactors had previously been achieved via the resistance produced when electrical current or radio waves were run through the plasma. The latest tokamaks employed a new technology, the neutral beam injector, which augmented heating by accelerating particles into the plasma. The device converts moving ions (which in transit would be influenced by the magnetic field such that they could only shallowly penetrate the plasma) into neutral atoms. These particles are then shot deep into the plasma, where they collide with electrons and ions, reionize, and become confined within the magnetic field, producing heat in the process.

Scientists have used tokamaks as research tools to explore a number of fusion problems, including energy confinement and the behavior of neutrons and helium ions in the fusion fuel reaction. Scientists consider a 50:50 deuterium–tritium (DT) combination the optimal fuel for two reasons. First, it has a fusion temperature of 100,000,000°C, the lowest of all fuel configurations. Second, because the chances of high-speed particles colliding and fusing in the superheated plasma increase with larger quantities of neutrons in the fuel mix, deuterium by itself is inefficient as a fuel because its nuclei have only one neutron. Deuterium must be combined with tritium, the nuclei of which carry two neutrons, for sustained, efficient fusion to occur in a tokamak. Though most of the energy in a DT reaction is produced by the release of fast neutrons, these particles are not affected by the tokamak magnetic field and escape with their plasma-heating energy. However, helium ions, along with DT ions, are subject to magnetic attraction, and their prolonged confinement is indispensable for plasma heating and, ultimately, ignition, the point at which a fusion reaction becomes self-sustaining.

The course of fusion experimentation has been governed by economic and environmental conditions. During the energy crisis of the 1970s, U.S. scientists focused on DT fuel dynamics in approaching the problem of ignition. The easing of the crisis in the 1980s and concern over the use of mildly radioactive tritium impelled researchers to examine the energy confinement question, resulting in experiments employing only deuterium. Not until the early 1990s were tests staged using DT. The JET produced the world's first significant controlled fusion power in 1991, using a 90:10 DT mixture to generate nearly two megawatts. In 1993, the TFTR produced the world's first 50:50 DT reaction, yielding three megawatts. The TFTR and the JET have since achieved peak outputs of 10.7 and 16 megawatts, respectively. However, tokamak reactors currently consume much more power than they yield. During a typical experiment, the JET requires up to 500 megawatts to supply its transformer, toroidal/poloidal field coils, and various heating appliances.

The TFTR ceased operation in 1997 and was replaced at Princeton in 1999 by the National Spherical Torus Experiment (NSTX), a small tokamak designed to explore plasma physics using only deuterium. Today, the JET is the only tokamak equipped to operate with tritium. The legacy of the tokamak reactor is an increased understanding of fusion theory and experience in reactor operations such that the scientific community currently believes ignition and "break-even" operations (where power input equals power output) may be possible with the planned International Thermonuclear Experimental Reactor (ITER). If built, ITER would be the largest and most advanced tokamak, with a projected output of 410 megawatts. The U.S. joined the European Union, Japan, and the Soviet Union as a formal ITER partner in 1988, but American participation has since fluctuated. In 1998, Congress directed the Department of Energy to withdraw from the project, although the U.S. rejoined talks with international participants in early 2003. Although negotiations continue today over the siting of ITER, with France and Japan as the leading candidates, the future of the project is not certain.

See also **Nuclear Reactor Materials; Nuclear Reactors, Weapons Material; Nuclear Reactors: Fast Breeders; Nuclear Reactors: Fusion, Early Designs; Nuclear Reactors: Thermal, Graphite-Moderated; Nuclear Reactors: Thermal, Water-Moderated**

MATTHEW EISLER

Further Reading

Bromberg, J.L. *Fusion: Science, Politics, and the Invention of a New Energy Source*. MIT Press, Cambridge, MA, 1982.
Butler, D. France and Japan lock horns in battle to host research reactor. *Nature*, 426, 4 December, 483, 2003.
Feng, K.M., Zhang, G.S. and Deng, M.G. Transmutation of minor actinides in a spherical torus tokamak fusion reactor, FDTR. *Fusion Eng. Design*, 63/64, 1, 127–133, 2002.
Fowler, T.K. *The Fusion Quest*. The Johns Hopkins University Press, Baltimore, 1997.
Gross, R.A. *Fusion Energy*. Wiley, New York, 1984.
Harms, A.A., et al., *Principles of Fusion Energy: An Introduction to Fusion Energy for Students of Science and Engineering*. World Scientific, River Edge, NJ, 2000.
Herman, R. *Fusion: The Search for Endless Energy*. Cambridge University Press, New York, 1990.
Mitrishkin, Y.V., Kurachi, K. and Kimura, H. Plasma multivariable robust control system design and simulation for a thermonuclear tokamak-reactor. *Int. J. Control*, 76, 13, 1358–1374, 2003.
Nishikawa, K. and Wakatani, M. *Plasma Physics: Basic Theory with Fusion Applications*. Springer Verlag, New York, 2000.
Souers, P.C. *Hydrogen Properties for Fusion Energy*. University of California Press, Berkeley, 1986.

Useful Websites

European Fusion Development Agreement-JET: http://www.jet.efda.org/pages/history-of-jet.html
International Thermonuclear Experimental Reactor: http://wwwofe.er.doe.gov/iter.html
Princeton Plasma Physics Laboratory: http://www.pppl.gov/news/pages/tftr_removal.html

Nuclear Reactors: Thermal, Graphite-Moderated

In a nuclear reactor, an element low on the atomic scale such as carbon or hydrogen is used to absorb kinetic energy to slow down naturally emitted neutrons from the radioactive fuel. In most power reactors, refined but unenriched natural uranium (^{238}U or uranium-238) is the preferred fuel over 99 percent of the time. When the neutrons move more slowly or at a "moderated" speed, the chances of collision between the neutrons and other uranium nuclei, leading to fission and a chain reaction, are increased. Reactor designs are often named for the type of moderator used.

The first reactors, including the experimental pile built in 1942 at Chicago during World War II and the early production reactors built in 1943 at Hanford in Washington state, used graphite as a moderator. Later reactors used water, heavy water, sodium, or other materials as moderators. In the U.S., almost all power reactors and all submarine and ship propulsion reactors relied on pressurized water systems or boiling water systems, first installed in the late 1950s. Acronyms for all these systems have become conventional, with the most common being the boiling water reactor (BWR), pressurized water reactor (PWR), and light water-cooled graphite-moderated reactor (LWGR).

The former Soviet Union developed three power reactor designs in the 1950s and 1960s. The *reactory bolshoi moshchnosti kanalynye* (RBMK) or "channelized large power reactor" was one of their most common and, equivalent to a LWGR, very similar to the first production reactors built in the U.S. at Hanford. A second Soviet type was the *vodo-vodyannoy energeticheskiy* reactor (VVER) or water-cooled and water-moderated reactor, rather similar to the American and European PWRs. The Soviets exported VVERs to several eastern European countries. The third Soviet type was a fast reactor designated "BN" as a breeder reactor.

Britain developed the Magnox reactor, a gas-cooled, graphite-moderated reactor, using the system both for plutonium production for weapons and for power production.

Accidents involving graphite reactors are particularly dangerous, since graphite is flammable. A release of radioactivity in 1957 at the British Windscale Reactor near Sellafield, Cumbria, a graphite production reactor, was not immediately disclosed. Even accidents with water-cooled reactors, such as that at Three Mile Island in Pennsylvania on March 28, 1979, cause national and international concern. However, far more serious was the Chernobyl fire of April 26 1986, in a 1000-megawatt rated RBMK graphite-moderated reactor. That fire spread radioactive contamination across not only the Ukraine but also much of eastern and northern Europe as well. As at Windscale, details of the Chernobyl accident were temporarily suppressed. Gas-cooled graphite reactors are prevented from burning by the fact that they are cooled with carbon dioxide. However, if oxygen-containing air leaks into the system and the cooling system fails, the graphite can ignite.

In the period from 1964 to 1987, a large LWGR used for both plutonium production and power

generation remained in operation on the Hanford reservation in the U.S. Designated "N" reactor, it was the only graphite-moderated reactor still in operation in the U.S. during the Chernobyl accident. After a safety review conducted by the National Academy of Sciences, N reactor was closed permanently.

The first British power reactor went into operation in 1956 at Calder Hall, in Seascale, Cumbria, at 50 megawatts electric (MWe). The early British design was called the Magnox type and was a gas-cooled reactor (GCR). Several of these GCRs were exported to France, Italy, Japan, and Spain. The Magnox reactor uses magnesium oxide as canning for the fuel slugs, and is cooled with carbon dioxide. The Magnox design had been worked out at Harwell, and early models were constructed at both Calder Hall and Chapelcross. A later design, first constructed at Hinkley Point B in Somerset is an advanced gas-cooled reactor (AGR). Whereas the GCRs operate at about 245°C, the AGRs operate at about 630°C. In some descriptions, the AGRs are known as high-temperature gas-cooled reactors.

Over the decades, Britain built more than 30 power reactors, providing nearly 20 gigawatts of power by the 1990s (about 12.5 gigawatts net). All are of the Magnox GCR type or the advanced gas-cooled reactor (AGR) type, except for one PWR, at Sizewell, Suffolk, built in the 1990s. The Magnox and AGR reactors differ in several aspects, but both are graphite-moderated.

At the beginning of the twenty-first century, the U.K. remained one of the largest producers of nuclear power in the world, ranking after the U.S., France, Japan, Germany, Russia, the Ukraine, and Canada. Nuclear power, nearly all of it from graphite-moderated reactors (see Table 1), made up slightly more than one fourth of the electricity supplied in the U.K.

See also **Nuclear Reactors: Fast Breeders; Nuclear Reactors: Fusion, Early Designs; Nuclear Reactors: Fusion, Tokomak; Nuclear Reactors: Thermal, Water-Moderated**

RODNEY CARLISLE

Further Reading

Carlisle, R., Ed. *Encyclopedia of the Atomic Age*. Facts on File, New York, 2001.

Carlisle, R. and Zenzen, J. *Supplying the Nuclear Arsenal*. Johns Hopkins University Press, Baltimore, 1996.

Cochrane, T.B., *et al. Making the Russian Bomb: From Stalin to Yeltsin*. Westview Press, Boulder, CO, 1995.

Hewlett, R.G. and Holl, J.M. *Atoms for Peace and War, 1953–1961*. University of California Press, Berkeley, 1989.

Patterson, W.C. *Nuclear Power*. Penguin, Harmondsworth, 1976

Rhodes, R. *The Making of the Atomic Bomb*. Simon & Schuster, New York, 1986.

Table 1 British Graphite Reactors.

Characteristic	Magnox or graphite core reactor	Advanced gas-cooled reactor (AGR)
Thermal output	800+ MWt	1400+MWt
Electrical output	50–475 megawatt electric	200–700 megawatt electric
Efficiency	c.33%	c. 40%
Moderator	Graphite	Graphite
Fuel	Natural uranium	Uranium oxide 2% enriched
Cladding	Clad in magnox alloy	Stainless steel
Coolant	Carbon dioxide gas	Carbon dioxide gas
Coolant temperature	245°C	634°C.
Vessel	Welded steel	Prestressed concrete
Hazard	Magnox alloy low melting point	Cladding has higher melting point

*Megawatt electric output as percentage of megawatt thermal output.

Nuclear Reactors: Thermal, Water-Moderated

Nuclear reactors are usually classified by their coolant and their moderators. The moderator is a material, low in the atomic scale, whose atomic nucleus has the effect of slowing down or moderating the speed of fast neutrons emitted during nuclear fission. By slowing the speed of neutrons, the moderator increases the chance of collision of neutrons with the nuclei of fissionable nuclear fuel atoms. The original reactor designed by Enrico Fermi during the Manhattan Project at Chicago, known as Chicago Pile One, or CP-1, was a graphite-moderated, air-cooled reactor. Many British and French nuclear reactors for the generation of electrical power use carbon in the form of graphite, and they are cooled with carbon dioxide gas. These types are known as Magnox reactors. However, the common designs for power generation developed in the U.S. used water both as coolant and as a moderator.

Water-cooled reactors fall into two large families. Heavy water reactors contain water in which the hydrogen atom is replaced with the hydrogen isotope deuterium. This type of reactor is manufactured for export by Canada. The pressurized heavy water reactor (PHWR) has been exported and installed in India, Romania, and elsewhere. The U.S. built five heavy water reactors at Savannah River, South Carolina, in the 1950s to serve as production reactors for the manufacture of plutonium and tritium for nuclear weapons. By the late 1980s, all the Savannah River production reactors had been closed. After some experimentation with graphite-moderated gas-cooled designs and with heavy-water moderation during the 1950s, the U.S. followed the "light water" path.

In the U.S., two reactor designs use regular or light water both as coolant and moderator. The light water reactor (LWR) has been designed in two forms. The smaller, pressurized water reactor (PWR), originally designed for use aboard submarines as a propulsion reactor for generating electricity to power the drive motors of the vessel, was the first adapted for use in the U.S. for power purposes. The first reactor of this type on land for power generation was installed at Shippingport, Pennsylvania, in 1957. The Shippingport reactor, installed under the supervision of the naval reactor chief, Admiral Hyman Rickover, was modeled on one designed for the propulsion of surface ships such as cruisers and aircraft carriers.

Most PWRs in the U.S. were manufactured by Westinghouse Corporation, while most boiling water reactors (BWRs) were manufactured by General Electric. In the pressurized water reactor type, the water is circulated through the reactor and super-heated, then piped through a steam generator where a separate water system is heated to produce steam. That steam is used to drive the turbine generators. In the BWR, a single loop takes water through the reactor, raising it to steam temperature that is then channeled through pipes and valves to the turbine. It is then condensed back to water to be pumped back into the reactor. In both systems, the steam is cooled by an independent system of outside water circulated through a characteristic cooling tower. For this reason, both types of light water reactors are often sited near a body of natural fresh water. The effluent from the cooling tower, although not radioactive in the slightest, is usually released back to the natural water supply at a somewhat elevated temperature. This "thermal pollution" has become a concern of some environmentalists, who decry its effect on local ecosystems.

By the beginning of the twenty-first century, all power reactors in the U.S. were light water types, with about two-thirds of the reactors as pressurized water reactors and about one-third following the boiling water reactor design. Advocates of gas-cooled designs in the U.S. argued that the inherent safety of the gas-cooled design, in which a runaway reaction automatically causes the reactor to close down, is superior to the water-cooled and moderated designs.

An advantage of the Canadian-designed PHWRs is that they operate on natural, unenriched uranium. One disadvantage is that the heavy water utilized in the reactors may be diverted for use in reactors to produce weapons-grade plutonium, as has been suspected in the case of the PHWR program in India.

One type of reactor made in the Soviet Union was the *vodo-vodyannoy energeticheskiy* reactor (VVER), or water-cooled and water-moderated reactor, equivalent to the American PWRs. The VVERs were built not only for use in the Soviet Union but also for export to satellite nations. The VVER-440, developed before 1970, was the most common. The later VVER-440s and a VVER-1000 that was developed in 1975 had added safety features.

The 1979 Three Mile Island accident, in which a Babcock and Wilcox-manufactured 906-megawatt-electrical PWR failed, was responsible more than any other factor for changing U.S. attitudes toward nuclear power. The accident began with a failure in the cooling system, resulting in an interruption in the flow of water to the steam generator. A series of errors, including valves

accidentally left in the closed position, unobserved warning lights, and a faulty pressure relief valve, all led to a spike in heat and reactivity.

Operators grew concerned that hydrogen produced by reaction of steam with the fuel cladding at high temperatures might explode when mixed with oxygen resulting from the breakup of water under radiation. The operators and outside emergency personnel disagreed over whether a hydrogen explosion could occur, but word of a possible explosion was released to the public. The Pennsylvania authorities ordered an evacuation of children and pregnant women from the immediate area surrounding the reactor. The after-accident appraisals were contradictory and depended to an extent on the orientation of the appraisal authors. Fortunately, there had been very little radioactive release. Even so, the damage to the reactor core was severe. The concrete containment vessel performed very well, and there were no injuries or deaths. The reactor was shut down and entombed in concrete. Despite problems, engineers concluded that the accident demonstrated that some existing safety systems worked very well.

Following the Three Mile Island accident, reactor orders in the U.S. greatly declined. Although new reactors have been built, and the U.S. remains the leader in electrical power generation from nuclear sources with about 100 power reactors in total, most specialists predict that the proportion of nuclear power production versus other sources will decrease in the U.S. in the coming decades.

See also **Nuclear Reactors: Fast Breeders; Nuclear Reactors: Fusion, Early Designs; Nuclear Reactors: Fusion, Later Designs (Tokomak); Nuclear Reactors: Thermal, Graphite-Moderated**

RODNEY P. CARLISLE

Further Reading

Balogh, B. *Chain Reaction: Expert Debate and Public Participation in American Commercial Power, 1945–1975*. Cambridge University Press, New York, 1991.

Bodansky, D. *Nuclear Energy: Principles, Practices, and Prospects*. American Institute of Physics, Woodbury, NY, 1996.

Cantelon, P. and Williams, R. *Crisis Contained: The Department of Energy at Three Mile Island*. University of Southern Illinois Press, Carbondale, 1984.

Duderstadt, J.J. and Kikuchi, C. *Nuclear Power: Technology on Trial*. University of Michigan Press, Ann Arbor, 1979.

Holl, J., Anders, R. and Buck, A. *The United States Civilian Nuclear Power Policy, 1954–1984: A Summary History*. Department of Energy, Washington D.C., 1986.

Kemeny, J.G. *Report of the President's Commission on the Accident at Three Mile Island*. Pergamon Press, New York, 1979.

Ragovin, M. *Three Mile Island: A Report to the Commissioners and to the Public*. Nuclear Regulatory Commission, Washington D.C., 1980.

Nuclear Reactors, Weapons Material

The first successful nuclear reactor, called an "atomic pile" because of its structure of graphite bricks, was completed and operational on December 2, 1942, in Chicago in the U.S. Although originally built to demonstrate a controlled nuclear reaction, the reactor was later dismantled and the depleted uranium removed in order to recover minute amounts of plutonium for use in a nuclear weapon. In effect, Chicago Pile-One (CP-1) was not only the world's first nuclear reactor but also the world's first reactor used to produce material for a nuclear weapon.

CP-1 and a few other experimental reactors built at Chicago were soon followed by a second generation of five reactors designed solely for large-scale production of plutonium, constructed by DuPont Corporation at Hanford, Washington. Dedicated to production of weapons material, they are designated "production reactors" to distinguish them from later power-generating reactors and ship or submarine propulsion reactors. Later production reactors also made the rare isotope of hydrogen, tritium, for use in weapons. Tritium causes a partial fusion reaction, boosting the neutron supply in a fission reaction.

Production reactors were first built in the U.S. in 1943 at Hanford to take advantage of the constant flow of cold water from the Columbia River for cooling the reactors. The isolation from centers of population made security easier and reduced the number of people who might suffer radiation exposure in case of an accident. The Hanford reactors consisted of 12-meter-high cores of graphite, through which horizontal tubes fed cold water from the river. Operators inserted slugs of tin-encased natural uranium into the tubes. The slugs were designed with fins to hold them in place in the tubes, allowing the cooling water to flow directly over the slugs (and later, through longitudinal holes, in the slugs themselves). Elements common to later reactors were included: moderated by graphite, controlled with cadmium rods, and cooled by natural water.

The uranium-238 in the natural uranium slugs was transformed by a several-step reaction into plutonium. The uranium-235 split into fission

fragments and generated two or three neutrons to sustain the reaction. Operators pushed the fuel slugs out of the back of the reactor and took them to a processing plant to separate plutonium. At the beginning of November 1944, the first reactor produced plutonium that was later used in the Trinity test at Alamogordo, New Mexico. The atomic weapon dropped on Nagasaki, Japan, in August 1945 also contained plutonium from the Hanford reactors.

Following World War II, an additional five reactors were built at Hanford. Responding to the Soviet test of a nuclear weapon in August 1949, and reacting to the outbreak of the Korean War in June 1950, the U.S. Congress funded a second production reactor complex at Savannah River, South Carolina. The five Savannah River Site reactors used heavy water as a moderator; that is, water based on the deuterium isotope of hydrogen. In the U.S., a few reactors were kept open throughout the Cold War, supplying both plutonium and the tritium isotope of hydrogen. The tritium was produced by the insertion of lithium deuteride pellets or targets in the reactors and bombarding them with neutrons.

All the nuclear-armed nations built reactors to make plutonium. The British facility at Calder Hall in Cumbria, the Soviet facilities at Chelyabinsk-40 and later at Krasnoyarsk-26 and at Tomsk-7, and the French Marcoule facility all housed production reactors. N reactor, built by the U.S. at Hanford in 1964, was a dual-purpose reactor designed both to serve as a production reactor for making plutonium and to generate electricity. Eventually other dual-purpose reactors included some of the Soviet reactors at Tomsk-7 and Krasnoyarsk-26, the British reactors at Calder Hall, and the French Marcoule complex. The U.S. Congress debated the wisdom of mixing the weapons material production function with electrical power production for the civilian power grid. In the U.S., the dilemma was resolved at N reactor by producing steam that was piped off site and used there to produce electricity. Apparently Russian experts disagreed over the propriety of the Russian dual-purpose program at Tomsk-7. The fact that some British dual-purpose reactors produced both electric power for commercial use and plutonium for weapons remained a matter of continuing controversy in the U.K. as well.

India maintains several production reactors, including Cirus and a larger one at Dhruva. One Indian reactor, Kakrapar II, opened in 1995 as a power reactor but could be used to generate material for weapons use. Pakistan maintains a small plutonium production reactor, rated at 50 megawatts at Khusab. Israel has a production reactor with a classified level of production in the Negev at Dimona. China is known to have two graphite-moderated production reactors. A 1000-megawatt reactor is located in Guangyuan, Sichuan province, and a smaller one, estimated at between 400 to 500 megawatts, is reputed to be at the Jiuquan Atomic Energy Complex at Subei, in Gansu province.

In the U.S., the last production reactors closed in 1988. In Russia, three production reactors remained in operation until the year 2000, including underground graphite-moderated reactors at Krasnoyarsk-26 (renamed Zheleznogorsk) and at Tomsk-7. In 1998, U.S. officials quietly announced that a power reactor in the Tennessee Valley Authority complex would be utilized for both power and weapons material production.

See also **Nuclear Fuels; Nuclear Reactor**

RODNEY P. CARLISLE

Further Reading

Carlisle, R., Ed. *Encyclopedia of the Atomic Age*. Facts on File, New York, 2001.

Carlisle, R. and Zenzen, J. *Supplying the Nuclear Arsenal*. Johns Hopkins University Press, Baltimore, 1996.

Cochrane, T.B., et al. *Making the Russian Bomb: From Stalin to Yeltsin*. Westview Press, Boulder, CO, 1995.

Gowing, M. *Britain and Atomic Energy 1939–1945*. Macmillan, London, 1964.

Gowing, M. Independence and Deterrence: Britain and Atomic Energy, 1945–1952. Macmillan, London, 1974.

Hewlett, R.G. and Anderson, O. *The New World, 1939–1946, vol.1: History of the United States Atomic Energy Commission*. Atomic Energy Commission, Washington D.C., 1972.

Hewlett, R.G. and Duncan, F. *Atomic Shield, 1947–1952, vol. 2: History of the United States Atomic Energy Commission*. Atomic Energy Commission, Washington D.C., 1972.

Hewlett, R.G. and Holl, J.M. *Atoms for Peace and War, 1953–1961*. University of California Press, Berkeley, 1989.

Holloway, D. *Stalin and the Bomb*. Yale University Press, New Haven, 1994.

Patterson, W.C. *Nuclear Power*. Cox & Wyman, London, 1976.

Rhodes, R. *The Making of the Atomic Bomb*. Simon & Schuster, New York, 1986.

Nuclear Waste Processing and Storage

Radioactive waste consists of liquid, solid, or gaseous materials containing high, medium, or low levels of radioactivity. There are two alternative policies underlying technologies for proces-

sing and storage of radioactive waste. If the level of radioactivity in the waste is low or medium, the materials are diluted in liquid carriers and then dispersed when the concentration of radioactivity is considered not hazardous for the environment. If the level of radioactivity in the waste is medium or high, then the radioactive waste is concentrated, contained, and isolated in proper repositories.

Low- and medium-level radioactive waste includes materials that have been contaminated in nuclear reactors during operation and that may contain fissile products or a low percentage of fuel. It also includes contaminated protective clothing and equipment. Some of the medium gaseous radioactive waste is due to fuel reprocessing. In this process, the dissolution of spent fuel in nitric acid causes the production of toxic gases containing radioactive isotopes such as krypton (^{85}Kr), iodine (^{129}I), carbon (^{14}C), and tritium (^{3}H). While tritium and carbon can be collected in liquid carriers, there is no solution currently available for krypton and iodine. These elements are therefore released into the atmosphere in diluted gaseous mixtures. Their containment consists of controlling the time, place and manner of dispersion.

Dispersion in the sea (or ocean dumping) was considered, in the past, a viable option for low-level radioactive waste. Waste was packed in steel or concrete drums and dumped at an average depth of 4000 meters in the sea. From 1946 to 1970, the U.S. dumped radioactive waste in 50 locations in the Pacific and Atlantic Oceans. The U.K. dumped radioactive waste in the Atlantic from 1949 to 1983. In the 1980s, rising concern was expressed for the marine biological environment and also for the possible irradiation of seafood with consequent hazard for human beings. In 1985 the United Nations imposed a moratorium on sea dumping of radioactive waste.

Processing and storage of high- and medium-level radioactive waste is considered the ''Achilles heel'' of the nuclear industry, as the technologies available have not yet provided viable solutions. Fission products such as iodine, technetium, neodymium, zirconium, molybdenum, and cesium produced by the fissile materials plutonium and uranium in nuclear reactors are by far the most toxic and radioactive elements known in nature, and their storage presents problems for the natural and human environment. Moreover, the problem faced by the scientific community is that a number of fission products are long-lived and will emit radiation and heat for a period varying from 10 to 100,000 years.

In the short term, high-level liquid and solid waste is stored in the proximity of the nuclear plants using water ponds. Subsequently, the high-level liquid waste is solidified through a process of vitrification with borosilicate glass. Vitrification was successfully achieved in France by the Commissariat À l'Energie Atomique (CEA) in 1978, and the Atelier de Vitrification de Marcule (AVM) was the first plant designed for this purpose. It represents an improvement in radioactive waste containment as the radioactive waste is chemically bonded into the atomic structure of an inert and stable solid. From 1988, the vitrification process was also adopted by Cogema in La Hague (France), by the U.K. Atomic Energy Agency at Windscale (Britain), and by the Vortec Corporation in the U.S. Borosilicate glass is also used for the production of canisters in which spent fuel and solid high waste is stored. Canisters will eventually be put into deep repositories capable of containing radioactivity and heat without causing major hazard to the environment, but no repositories of this kind have yet been accomplished.

Between the 1950s and the 1980s, several sites were used for short-term waste using shallow burial in trenches, tumuli, tunnels, or concrete bunkers. The oldest site of this type is at Savannah River (Idaho Desert, U.S.), which stores the contaminated material from Los Alamos National Laboratory in New Mexico. The U.K.'s short-term waste repository is located at Drigg in Cumbria. More modern facilities for short-term repositories use concrete pipes and concrete bases sunk vertically to prevent leakage as at Hanford in Britain, Tokai in Japan, and La Manche in France.

Since 1960, three types of geological sites have been considered for long-term radioactive waste disposal: clay-rich rocks, hard crystalline igneous and metamorphic rocks (granite, gabbro, and basalt), and evaporites (bedded salts or salt domes). In 1960, the Project Salt Vault for the creation of a facility in the former Lyons salt mine in Kansas (U.S.) was discussed, but in 1971 the project was finally dropped. A new project for repositories in deep salt beds was considered in 1988, but it never led to final construction. From 1982 onward, the scientific community stressed that deep geological disposal should be considered the final target for any high-level waste strategy. It also provided details of how containment should be pursued. Facilities should be provided with barriers capable of dividing the depository from the surrounding area. Repositories should contain the canisters, ventilate them with air, and survey them with CCTV systems. They should be built in

concrete and manufactured according to a geological analysis of the rocks, the water, and seismic activity. At the moment the only repository that may be used for long-term disposal is the undersea plant of Sipra in Sweden. Various projects for long-term disposal have not been developed because they have met severe public protests by local communities unwilling to accept radioactive waste in the proximity of their living areas (as highlighted by the acronym NIMBY for Not-In-My-Back-Yard).

In recent years, several scientific establishments have considered a different approach to long-term radioactive waste, focusing on the destruction of fission products. In 1997, the European Organisation for Nuclear Research presented a project for an energy amplifier capable of incinerating waste through fast neutron bombardment. The project aimed to substitute in 29 years the geological disposal of the high waste produced by nine nuclear reactors operating in Spain. In 2000, the Nuclear Energy Agency (OECD-NEA) published the results of an international meeting on the process of partitioning and transmutation (P&T). The goal is to shorten the life of some fission products through incineration in nuclear reactors or treatment in nuclear accelerators. P&T research programs have been developed in Japan (Omega, 1988) and in France (Spin, 1991).

See also **Nuclear Reactors, Weapons Material**

SIMONE TURCHETTI

Further Reading

Annex A: The Spin Programme—Assets and Prospects, in Assessment Report on Actinide and Fission Product Partitioning and Transmutation. Sixth International Information Exchange Meeting, Madrid, Spain, 11–13 December, 2000.

Annex B: The Omega Programme—Partitioning and Transmutation R&D Programme of Japan, in Assessment Report on Actinide and Fission Product Partitioning and Transmutation. Sixth International Information Exchange Meeting, Madrid, Spain, 11–13 December 2000.

Blowers, A., Lowry, D. and Solomon, B.D. The International Politics of Nuclear Waste. Macmillan, London, 1991.

Mounfield, P. World Nuclear Power. Routledge, London, 1991.

Rubbia C., Buono S., Kadi Y. and Rubio J.A. *Fast Neutron Incineration in the Energy Amplifier as Alternative to Geological Storage: the Case of Spain*. European Organisation for Nuclear Research, Geneva, 1997.

Ursu, I. *Physics and Technology of Nuclear Materials*. Pergamon Press, Oxford, 1985.

O

Oil from Coal Process

The twentieth-century coal-to-petroleum or synthetic fuel industry evolved in three stages:

1. Invention and early development of Bergius coal liquefaction (hydrogenation) and Fischer–Tropsch (F–T) gas synthesis from 1910 to 1926.
2. Germany's industrialization of the Bergius and F–T processes from 1927 to 1945.
3. Global transfer of the German technology to Britain, France, Japan, Canada, the U.S., and other nations from the 1930s to the 1990s.

Petroleum had become essential to the economies of industrialized nations by the 1920s. The mass production of automobiles, the introduction of airplanes and petroleum-powered ships, and the recognition of petroleum's high energy content compared to wood and coal, required a shift from solid to liquid fuels as a major energy source. Industrialized nations responded in different ways. Germany, Britain, Canada, France, Japan, Italy, and other nations, having little or no domestic petroleum, continued to import it. Germany, Japan, and Italy also acquired by force the petroleum resources of other nations during their 1930s–1940s World War II occupations in Europe and the Far East. In addition to sources of naturally occurring petroleum, Germany, Britain, France, and Canada in the 1920s–1940s synthesized petroleum from their domestic coal or bitumen resources, and during the 1930s–1940s war years Germany and Japan synthesized petroleum from the coal resources they seized from occupied nations. A much more favorable energy situation existed in the U.S., and it experienced few problems in making an energy shift from solid to liquid fuels because it possessed large resources of both petroleum and coal.

Germany was the first of the industrialized nations to synthesize petroleum when Friedrich Bergius in Rheinau-Mannheim in 1913 and Franz Fischer and Hans Tropsch at the Kaiser Wilhelm Institute for Coal Research (KWI) in Mülheim in 1926 invented processes for converting coal to petroleum. Bergius crushed and dissolved a coal containing less than 85 percent carbon in a heavy oil to form a paste. He reacted the coal–oil paste with hydrogen gas at high pressure (200 atmospheres at 400°C) and obtained petroleum-like liquids. Bergius sold his patents to BASF in July 1925, and from 1925 to 1930 Matthias Pier at BASF (IG Farben in December 1925) made major advancements that significantly improved product yield and quality. Pier developed sulfur-resistant catalysts, such as tungsten sulfide (WS_2), and separated the conversion into two stages, a liquid stage and a vapor stage. Fischer and Tropsch reacted coal with steam to give a gaseous mixture of carbon monoxide and hydrogen and then converted the mixture at low pressure (1–10 atmospheres at 180–200°C) to petroleum-like liquids. Fischer and his co-workers in the 1920s–1930s developed the cobalt catalysts that were critical to the F–T's success, and in 1934 Ruhrchemie acquired the patent rights to the synthesis. These pioneering researches enabled Germany to develop a technologically successful synthetic fuel industry that grew from a single commercial-size coal liquefaction plant in 1927 to twelve coal liquefaction and nine F–T commercial-size plants that in 1944

reached a peak production of 26 million barrels of synthetic fuel.

Britain's synthetic fuel program evolved from post-World War I laboratory and pilot-plant studies that began at the University of Birmingham in 1920 on F–T synthesis and in 1923 on coal liquefaction. The Fuel Research Station in East Greenwich also began research on coal liquefaction in 1923, and the program reached its zenith in 1935 when Imperial Chemical Industries (ICI) constructed a coal liquefaction plant at Billingham that had the capacity to synthesize annually 1.28 million barrels of petroleum. British research and development matched Germany's, but because of liquefaction's high cost and the government's decision to rely on petroleum imports rather than price supports for an expanded domestic industry, Billingham remained the only British commercial-size synthetic fuel plant. F–T synthesis in the 1930s–1940s never advanced beyond the construction of four small experimental plants: Birmingham, the Fuel Research Station's two plants that operated from 1935 to 1939, and Synthetic Oils Ltd. near Glasgow.

Britain and Germany had the most successful synthetic fuel programs. The others were either smaller-scale operations such as France's three demonstration plants (two coal liquefaction and one F–T), Canada's bitumen liquefaction pilot plants, and Italy's two crude petroleum hydrogenating (refining) plants; or technological failures such as Japan's five commercial-size plants (two coal liquefaction and three F–Ts) that produced only about 360,000 barrels of liquid fuel during the World War II years.

The U.S. Bureau of Mines had begun small-scale research on F–T synthesis in 1927 and coal liquefaction in 1936, but did no serious work on them until the government expressed considerable concern about the country's rapidly increasing petroleum consumption in the immediate post-World War II years. At that time the Bureau began a demonstration program, and from 1949 to 1953 when government funding ended, it operated a small 200–300 barrel per day coal liquefaction plant and a smaller 50 barrel per day F–T plant at Louisiana, Missouri. In addition to the Bureau's program, American industrialists constructed four synthetic fuel plants in the late 1940s and mid-1950s, none of which achieved full capacity before shutdown in the 1950s for economic and technical reasons. Three were F–T plants located in Garden City, Kansas; Brownsville, Texas; and Liberty, Pennsylvania. The fourth plant was a coal liquefaction plant in Institute, West Virginia.

Following the plant shutdowns in the U.S. and until the global energy crises of 1970–1974 and 1979–1981, all major synthetic fuel research and development ceased except for the construction in 1955 of the South African Coal, Oil, and Gas Corporation's (SASOL) F–T plant in Sasolburg, south of Johannesburg. South Africa's desire for energy independence and the low quality of its coal dictated the choice of F–T synthesis rather than coal liquefaction. Its Johannesburg plant remained the only operational commercial-size synthetic fuel plant until the 1970s energy crises and South Africa's concern about hostile world reaction to its apartheid policy prompted SASOL to construct two more F–T plants in 1973 and 1976 in Secunda.

The 1970s energy crises also revitalized synthetic fuel research and development in the U.S. and Germany and led to joint government–industry programs that quickly disappeared once the crises had passed. Gulf Oil, Atlantic Richfield, and Exxon in the U.S., Saarbergwerke AG in Saarbrüken, Ruhrkohle AG in Essen, and Veba Chemie in Gelsenkirchen, Germany, constructed F–T and coal liquefaction pilot plants in the 1970s and early 1980s only to end their operation with the collapse of petroleum prices a few years later.

In the mid-1990s two developments triggered another synthetic fuel revival in the U.S.:

1. Petroleum imports again reached 50 percent of total consumption or what they were during the 1973–1974 Arab petroleum embargo.
2. An abundance of natural gas, equivalent to 800,000 million barrels of petroleum, but largely inaccessible by pipeline, existed.

Syntroleum in Tulsa, Oklahoma; Exxon in Baytown, Texas; and Atlantic Richfield in Plano, Texas, developed modified F–T syntheses that produced liquid fuels from natural gas and thereby offered a way of reducing the U.S.'s dependence on petroleum imports. The Department of Energy (DOE) at its Pittsburgh Energy Technology Center through the 1980s–1990s also continued small-scale research on improved versions of coal liquefaction. DOE pointed out that global coal reserves greatly exceeded petroleum reserves, anywhere from 5 to 24 times, and that it expected petroleum reserves to decline significantly in 2010–2030. Syntroleum, Shell in Malaysia, and SASOL and Chevron in Qatar have continued F–T research, whereas DOE switched its coal liquefaction research to *standby*. The only ongoing coal

liquefaction research is a pilot plant study by Hydrocarbon Technologies Inc. in Lawrenceville, New Jersey, now Headwaters Inc. in Draper, Utah.

A combination of four factors, therefore, has led industrialized nations at various times during the twentieth century to conclude that synthetic fuel could contribute to their growing liquid fuel requirements:

1. The shift from solid to liquid fuel as a major energy source.
2. The invention of the Bergius and F–T coal-to-petroleum conversion or synthetic fuel processes.
3. Recognition that global petroleum reserves were finite and much less than global coal reserves and that petroleum's days as a plentiful energy source were limited.
4. The desire for energy independence.

With the exception of South Africa's three F–T plants, the synthetic fuel industry, like most alternative energies, has endured a series of fits and starts that has plagued its history. The historical record has demonstrated that after nearly 90 years of research and development, synthetic liquid fuel has not emerged as an important alternative energy source. Despite the technological success of synthesizing petroleum from coal, its lack of progress and cyclical history are the result of a lack of government and industry interest in making a firm and a long-term commitment to synthetic fuel research and development. The synthetic fuel industry experienced intermittent periods of intense activity internationally in times of crises, only to face quick dismissal as unnecessary or uneconomical upon disappearance of the crises. Even its argument that synthetic liquid fuels are much cleaner burning than coal, and if substituted for coal they would reduce the emissions that have contributed to acid rain formation, greenhouse effect, and to an overall deterioration of air quality has failed to silence its critics. The hope of transforming its accomplishments at the demonstration stage into commercial-size production has not yet materialized.

See also **Feedstocks**

Anthony N. Stranges

Further Reading

Much of the information on the German, Japanese, British, and American synthetic fuel programs has come from the 300,000 pages of documents and the 305 microfilm reels that the Technical Oil Mission and other Allied investigative teams in Europe collected on the German synthetic fuel program, and the 185 microfilmed reports that the U.S. Naval Technical Mission to Japan compiled. The reports on the Japanese synthetic fuel industry are the same as those in the National Diet Library in Tokyo and in the National Archives in Washington. Complete sets of both microfilm collections and documents are in the Texas A&M University Archives.

Donath, E. Hydrogenation of coal and tar, in *Chemistry of Coal Utilization,* suppl. vol., Lowry, H.H., Ed. Wiley, New York, 1963.

Report on the Petroleum and Synthetic Oil Industry of Germany. The Ministry of Fuel and Power, His Majesty's Stationery Office, London, 1947.

Storch, H.H. Golumbic, N. and Anderson, R. *The Fischer–Tropsch and Related Syntheses*. Wiley, New York, 1951.

Stranges, A.N. Canada's mines branch and its synthetic fuel program for energy independence. *Technol. Cult.*, 32, 521–554, 1991.

Stranges, A.N. From Birmingham to Billingham: high-pressure coal hydrogenation in Great Britian. *Technol. Cult.*, 20, 726–757, 1985.

Stranges, A.N. Germany's synthetic fuel industry 1927–45, in *The German Chemical Industry in the Twentieth Century*, Lesch, J.E., Ed. Kluwer, Dordrecht, 2000.

Stranges, A.N. Synthetic fuel production in Japan: a case study in technological failure. *Ann. Sci.*, 50, 229–265, 1993.

Stranges, A.N. Synthetic petroleum from coal hydrogenation: its history and present state of development in the United States. *J. Chem. Ed.*, 60, 617–625, 1983.

Stranges, A.N. The US Bureau of Mines' synthetic fuel programme. *Ann. Sci.*, 54, 29–68, 1997.

Useful Websites

Fischer-Tropsch Archive: http://www.fischer-tropsch.org

Oil Rigs

Although historical accounts exist that describe oil and natural gas drilling techniques in ancient Mesopotamia and China, modern oil rig drilling has its roots primarily in salt-boring technology. By AD 350, China was constructing salt drilling wells that ran as deep as 900 meters into the ground. In the nineteenth century, Europe and the U.S. began importing this salt drilling technology from China. George Bissell, an American entrepreneur, realized that salt-boring techniques could be applied to the drilling for oil. Bissell and other investors hired Edwin Drake to construct and oversee rigs designed for oil drilling. Their venture proved successful when on 27 August 1859, Drake struck oil in Titusville, Pennsylvania.

Drake utilized cable tool drilling, or percussion drilling, which consists of a cable that raises and drops a heavy metal bit capped with a chisel end into the ground. The bit punches a hole into the

earth by breaking through rock with regularly repeated blows. Although cable tool drilling remains in use today in shallow, low-pressure oil and gas wells, rotary drilling has become the industry standard.

The transition from cable tool drilling to rotary drilling was the most significant advancement in rig-drilling technology during the twentieth century. Anthony Lucas and Patillo Higgins popularized rotary drilling when they used it in 1901 on their Spindeltop well located near Beaumont, Texas. Rotary drilling involves the use of a sharp, rotating drill bit which can cut through the hardest rock and dig deeper and higher pressured wells than is possible with cable tool drilling. Rotary drilling technology is composed of four basic components:

1. The prime mover, which provides power to the drilling rig.
2. The hoisting equipment, which includes the derrick that raises and lowers necessary equipment into the well hole.
3. The circulating system, which controls the well's pressure, cools and lubricates the drill bits, removes debris, and helps reduce the chance for a blowout.
4. The rotating equipment, which enables the rotation of the drill bit.

Offshore drilling dates back to the mid-nineteenth century when American T.F. Rowland was granted a patent for his design of a four-legged offshore drilling rig. However, it was not until the 1940s that North American construction of offshore wells began to flourish. Although onshore and offshore drilling technologies are similar, offshore drilling poses unique challenges. For example, onshore drilling uses the ground as a platform, but in the offshore environment, the floor can be located thousands of feet below sea level thus requiring the construction of an artificial drilling platform.

Offshore drilling rigs are located on either moveable or permanent sites. Moveable rigs are designed for mobility, which enable exploratory drilling opportunities in numerous places. The drilling barge, the jack-up rig, and the submersible rig are designed for use in shallow water. The semisubmersible rig and the drillship are able to withstand the often harsher conditions present in deeper water.

Both the drilling barge and the jack-up rig must be towed to their drilling site in shallow water. The drilling barge is a large, floating platform that drills in inland, calm water. The jack-up rig also drills in calm water. It is equipped with "legs" that can be lowered to the sea bottom, which, unlike the drilling barge, enables its working platform to remain above the water's surface.

Like jack-up rigs, submersible rigs are designed for direct contact with the lake or ocean floor. These rigs contain a platform that has two vertically connected hulls. The crew and the drilling platform reside in the top hull. The lower hull is equipped with a circulatory air system. When the hull is filled with air, the entire rig becomes buoyant which provides it with the ability to move from site to site. When the air is released, the rig drops to the floor. However jack-up rigs are not sturdy enough for use in deep water.

The most common offshore drilling rig is the semisubmersible rig. The semisubmersible rig is similar to the submersible rig in that it also possesses a lower hull that can inflate and deflate. However, when the semisubmersible rig's lower hull releases its air, the rig floats above the drill site instead of submerging to the floor. The lower hull is then filled with water, which provides the rig with the needed stability to drill in deep water. In addition, these rigs are held in place by anchors that weigh up to 10 tons each. This design provides semisubmersible rigs with the necessary reliability for drilling in offshore turbulent waters.

Drillships are used for drilling in very deep waters. These ships carry a drilling platform and a derrick, and are designed with a hole, called a "moonpool," which extends through the ship's hull. Drilling is done through the moonpool with the assistance of sensors and satellite positioning systems that ensure that the drillship is directly above the desired site.

Once a moveable drilling rig locates an oil or natural gas reservoir, a permanent platform can be built over the site to continue the drilling. These offshore rigs are among the largest man-made artifacts on the planet. Their height ranges in size from twice that of the Hoover Dam upward to that of the Empire State Building. The type of permanent platforms utilized for drilling is dependent on the depth range of the water. Fixed platforms are used in waters measuring up to around 450 meters. Compliant towers are used for drilling in sites where the floor is 450 to 900 meters below sea level. Seastar platforms are used where the drilling sites are located at 150 to 1060 meters below sea level. For deeper water levels, floating production systems can be used at up to 1.8 kilometers below sea level, tension leg platforms at up to 2.1 kilometers below sea level, and the SPAR system at up to 3 kilometers below sea level.

Although petroleum and natural gas provides much of the globe with its energy, drilling technology has proven controversial. Citizen and public interest groups have voiced their concerns about the negative impact drilling has on the environment. In addition, countries are increasingly confronting the challenges posed by decommissioned offshore rigs.

See also **Prospecting, Minerals**

Julia Chenot Goodfox

Further Reading

Gatlin, C. *Petroleum Engineering: Drilling and Well Completions*. Prentice Hall, Englewood Cliffs, NJ, 1960.
Hall, R. S., Ed. *Drilling and Producing Offshore*. PennWell Books, Tulsa, 1984.
Hyne, N.J. *Nontechnical Guide to Petroleum Geology, Exploration, Drilling and Production*, 2nd edn. PennWell Books, Tulsa, 2001.
Uren, L. *Petroleum Production Engineering*, 4th edn. McGraw-Hill, New York, 1956.
Yergin, D.Y. *The Prize: The Epic Quest for Oil, Money and Power*. Free Press, New York, 1991.

Useful Websites

American Petroleum Institute: http://api-ec.api.org
Greenpeace: http://archive.greenpeace.org
Natural Gas Supply Association: http://www.naturalgas.org

Ophthalmology

Ophthalmology is the medical specialty concerned with eye diseases and, in contrast to optometry, it deals with the pathologies of anatomy that require surgery or medical treatment. Ophthalmologists use more technology because they operate on the eye, but they share many of the technologies used to test and assess vision. Training for ophthalmology requires an MD degree followed by four years of specialization.

Prior to the nineteenth century, ophthalmology was included in the ear, nose, and throat (ENT) specialty. The first professor of ophthalmology in Europe was Joseph Baer (1812). Albrecht von Graefe founded the first archive for ophthalmological instruments and established the Ophthalmological Society in Heidelberg. In 1929 that name was changed to The German Ophthalmological Society. In the U.S., the American Board of Ophthalmology was established in 1916 to certify practitioners and maintain standards for the profession. At the time that the Canadian Ophthalmological Society was formed in 1937, ophthalmologists were part of the Eye, Ear, Nose and Throat (EENT) section of the Canadian Medical Association. In the United Kingdom, The European Ophthalmic Pathology Club was officially established at the Royal College of Surgeons in 1962. Worldwide there are over 300 ophthalmological societies, networks and associations.

Astronomers and physicists contributed to the early development of ophthalmology with their knowledge of lenses and the behavior of light. Famed astronomer Johannes Kepler wrote about the role of the retina and the lens in vision. Most surgery was for removal of foreign objects in the eye or cataract couching (or extraction), a fearsome procedure prior to the eighteenth century. Although this procedure to remove cataracts, a clouding of the lens that blocked the passage of light and caused poor vision or blindness, had been known since ancient times (as described in Sanskrit documents), it was dangerous. Both Bach and Handel died from postsurgical infections after successful removal of their cataracts in the eighteenth century. At that time, only the couching knife and needle, crude surgical instruments, were available. The technique involved perforating the sclera and pushing the lens backward with a blunt instrument. Jacques Daviel, born in Normandy, France, educated in Rouen and professor at Hotel Dieu in Marseille (1693–1752), considered the father of ophthalmology, changed the technique to perforate the cornea and remove the whole lens, which was known as extracapsular extraction.

Until the invention of the ophthalmoscope in 1851 by Hermann von Helmholtz, the terms glaucoma (a condition of increased pressure in the eye that can damage the optic disk and cause loss of vision) and cataract were often used interchangeably because no one had been able to see inside the eye of a living person. The first ophthalmoscope used a concave mirror mounted on a handle. In the center of the mirror was a hole through which light was shone into the patient's eye. The viewer saw the reflected image of the retina, the most posterior structure of the eye. It is there that abnormalities of the fundus such as glaucoma could be visualized. Cataracts, however, are found on the lens, a more anterior structure. The physician could finally distinguish the two conditions using the ophthalmoscope

Nineteenth century ophthalmic instruments were elegant-looking brass or copper polished metal materials mounted on solid ivory, ebony, or other kinds of wooden bases. They varied in their design and were operated by reflected light sources or natural light. In the early 1900s, the basic diagnostic

tools were the slit-lamp, ophthalmoscope, and retinoscope, which were already in use by both optometrists and ophthalmologists. The tonometer was also used to measure the intraocular pressure of the eye fluid, significant for the diagnosis of glaucoma. As the advent of plastics and other metals flooded the technology market in the 1940s, instruments became lighter, smaller, and more easily operated. The use of a battery-operated light source allowed the clinician more mobility. This was useful when examining the elderly or very young children who were unable to follow directions and position themselves correctly during examinations. The late twentieth century tonometer available was hand-held, used a 1.5-volt battery, and could be manipulated with only one hand.

In the 1950s a new type of light using fused quartz instead of a glass bulb was developed by General Electric. By the early 1960s, a tungsten halogen lamp was developed, and by the 1990s, a very bright halogen light source was used for ophthalmic instruments.

The invention of the slit lamp and its subsequent improvement through the final decades of the nineteenth century and first two decades of the twentieth century made it possible to study the cornea and anterior segment of the living human eye in a thorough and meaningful manner. Alfred Vogt in Switzerland and Allvar Gullstrand in Sweden are credited with its development. In the early 1930s, Vogt published *Lehrbuch und Atlas der Spaltlampenmikroskopie des Leibenden Auges,* which provided the foundation for a new branch of ophthalmology called slit-lamp biomicroscopy.

With the use of topical anesthetics, originally by Carl Koller for trachoma and iritis, and the advent of antibiotics in the mid-twentieth century, great strides were made in eye surgery. Not only could cataracts be removed safely but all kinds of microsurgical procedures were introduced.

After lens removal in cataract surgery, the patient had to wear glasses with extremely thick and heavy lenses and often had problems with depth perception and peripheral vision. The application of a contact lens solved this problem. However, since cataract is a condition more common in elderly patients, there was a need for a more permanent lens, one that would not require the fine motor manipulation required for frequent removal, replacement, and cleaning. During World War II, a serendipitous observation among pilots whose eyes were injured but not damaged by plastic windshield fragments led to the development of the intraocular lens implant. In the ensuing years, a physician to Royal Air Force pilots also observed that pilots who had sustained the intrusion of slivers of cockpit glass did not appear to have a foreign body reaction. He hypothesized that a small lens could be fashioned from similar material. This initial idea resulted in a successful plastic lens design for an intraocular lens implant to use in cataract surgery.

Radial keratotomy was a procedure developed in Japan in the 1940s and enthusiastically continued and perfected in Russia in the late 1950s. In the mid-1970s, Svyatoslav Fyodorov (Moscow) began performing refractive keratotomy with rather primitive instruments: a small razor blade in a holder and a gauge to check the depth of the incision. This revolutionary surgery changed the curvature of the cornea so that a person with myopia (nearsightedness) no longer needed glasses. Ophthalmologists from the U.S. traveled to Russia to learn the technique and used it until the 1990s, eventually replacing the blade with a laser. At the end of the twentieth century, *lasik* surgery to correct vision was performed at outpatient centers and valued because of its convenience and cosmetic result.

Many tests for vision and pathologies of the eye were developed during the twentieth century. The Amsler grid was the simplest and least expensive. It consists of a rectangle divided into small squares with a dot in the middle. The patient stares at the dot and then notes if any of the lines appear distorted. If so, the distortion indicates an underlying problem with the retina, and the patient would seek a thorough ophthalmological examination. More expensive and complex tests such as angiography, computed tomography (CT) scan, and magnetic resonance imagery (MRI) scan require the use of contrast materials and hospital equipment.

Digital technology is now used in cameras such as the fundus camera, which connects to the slit lamp or a combination of imaging devices used in tandem. The image is displayed on a monitor and can be manipulated, stored, and added to a database. A portable slit lamp is now available that is cordless, has a rechargeable battery, uses a halogen light source, and can support a digital video camera (made by Kowa Optimed).

The rise in ophthalmic technology during the twentieth century became so specialized that new educational programs were started in the U.S. to accommodate the rising need for workers within the field. One program at Eastern Virginia Medical School/Old Dominion University opened in 1985 and offered the following subspecialties: ophthalmic photography, ophthalmic ultrasonography, contact lens, ophthalmic surgical technology, electrophysiology, and low vision optics.

See also **Lasers, Applications; Optometry**

LANA THOMPSON

Further Reading

Garrison, F.H. *An Introduction to the History of Medicine*. W.B. Saunders, Philadelphia, 1929.

Lyons, A. and Petrucelli, J. *Medicine: An Illustrated History*. Harry Abrams, New York, 1987.

Magner, L. *A History of Medicine*. Marcel Dekker, New York, 1992.

Medow, N. and Ravin, R. Slit lamp provided closer glimpse at eye. *Ophthalmology Times*. November 1, 1999.

Useful Websites/Further Information

http://www.eyemdlink.com
http://www.eyesite.com
Haag-Streit USA. 5500 Courseview Drive, Mason, OH 45040-2303
Keeler Instruments Inc., 456 Parkway, Broomall, PA 19008
Kowa Optimed, Inc. 20001 S. Vermont Avenue. Torrance, CA 90502
National Optronics, 100 Avon Street, Charlottesville, VA 22902

For information about ophthalmologic technician programs, contact: Lions Center for Sight, 600 Gresham Drive, Norfolk, Virginia 23507,(757) 668-3747, E-mail: optech@-visi.net

Optical Amplifiers

The history of optical telecommunications has been the story of the overcoming of successive technical roadblocks. With the removal of each roadblock came a leap in possible transmission distances or data transmission rates.

The concept of transmitting light down ''glassy lightpipes'' (optical fibers) having taken root in some minds in the late 1960s, the obvious first objective was to reduce the attenuation (the weakening of the light beam with distance). Once this was achieved, and power loss of less than a factor of 10 in 20 kilometers was feasible, the next fundamental limitation was dispersion—the smearing of the pulses of light. A major improvement in dispersion was achieved with the development of single-mode fibers, which replaced multimode fibers and their intermodal dispersion problems. With this achievement, researchers once more focused on attenuation, since the new high data rate communications were limited to fiber spans of 40 to 60 kilometers between expensive electronic regeneration equipment. Long-distance telecommunications companies, which routinely carried traffic much further than 100 kilometers, saw the economic advantages of systems capable of transmitting high optical data rates for hundreds of kilometers. Proponents of optical communications promoted a vision of WDM (wavelength division multiplexing) that made this value even greater, but also made critical the development of a transparent signal regeneration technology, since no scenario in which wavelengths were separated at each regenerator and then recombined made business sense.

While there remained a prospect for the further reduction of fiber attenuation, theoretical limits seemed to indicate that only about a factor of 2 in reach extension could be achieved. The required reach extension was 10 times. Fortunately there was an alternative approach—optical amplification—in which the power of a light beam was increased (amplified) directly. Using stimulated emission, the same physics that underlies lasers, it was theoretically possible to prepare an ''excited'' medium through which light would pass, being amplified as it went. This amplification would preserve the phase relationships of the photons, so that any signal imposed on the light stream would be preserved. The theoretical cost is an admixture of random, or spontaneous, emission—background light that causes noise on the detector. This signal degradation is more than offset by the massive reduction in shot noise achieved by maintaining the signal power.

There were at least two obvious techniques for achieving optical signal gain, Raman amplification and amplification in semiconductor materials (semiconductor optical amplifiers, SOAs). Raman amplification occurs in all materials with its details depending on the material, and requires an intense ''pump'' beam of light at a wavelength shorter than that of the signal to drive the process, but will provide gain at a wavelength that can be controlled through the pump wavelength. SOAs were a clear extension of semiconductor lasers, and would share their advantages of bandgap engineering, allowing the material to be designed for operation in the required wavelength range, and direct electrical pumping. Both approaches would naturally lend themselves to a guided wave approach, Raman in optical fibers and SOAs in the waveguide structures already well developed for semiconductor lasers.

In fact, the practical realization of optical amplification came from quite a different direction, from a technology that did not offer the benefits of flexible operating wavelength. Researchers showed that to get to the threshold for Raman amplification required pump powers beyond the reach of cost-effective and practical implementations of the day. Engineers also failed to overcome the limita-

tions in SOA attributes, ranging from functional impairments like limited bandwidth and extra underlying attenuation to more fundamental processes such as cross-talk (interference) between channels of different wavelengths, and intersymbol interference (ISI) within a channel.

Optical fiber made of very dry silica has a wavelength "window" of low attenuation, roughly from 1500 to 1600 nanometers in the near-infrared range (known for historical reasons as the third telecommunications window). While this window would have difficulties with intramodal dispersion at very high bit rates, it was otherwise an attractive region in which to operate. Therefore, any amplification scheme that could work in this window would be technically advantaged, even if it were not applicable to other wavelength bands. At the end of the 1980s, the amplification technology that leaped to prominence was the EDFA, the erbium-doped fiber amplifier. In some ways derived from ideas of the early 1960s, the approach was similar to well-understood laser systems. The fundamental problem was to fabricate an optical fiber with the light-guiding core doped with a small quantity of a rare-earth element. The rare-earth elements have rich spectroscopy in the visible and near-infrared region, and neodymium and erbium turned out to be particularly interesting to telecommunications. Once researchers in fiber fabrication invented processes to manufacture the fiber, erbium ions in a silica–germania host glass proved to be an almost ideal medium to provide gain in the short wavelength end of the third telecommunications window. The first EDFA was realized by researchers at the University of Southampton in the U.K. in 1986. These researchers and their counterparts worldwide quickly demonstrated effective performance, with gains of up to 30 decibels (a power increase of 1000 times) for relatively low pump powers, which scaled with the signal power. The pump light could be derived from existing fiber-coupled laser diodes, and there was the prospect of using different pump wavelengths with even more favorable properties if the laser diode technology could be developed.

In retrospect, we can see that the development of EDFAs in the late 1980s, with early commercialization at the end of the decade and in the early 1990s, set the stage for over a decade of continuous rapid improvement in optical telecommunications systems and drove rapid and major enhancements in semiconductor laser technology, dispersion compensation techniques and optical componentry. From the very first, practical EDFAs required pump/signal multiplexers, diode pumps and fiber-pigtailed isolators, to prevent these extremely high-gain devices from being turned into lasers by inevitable stray back-reflections. The EDFA combined almost immediate economic dominance by its ability to replace electrical regenerators, and its great potential value, which was clear to many, could be unlocked by developments in sister technologies. Therefore as pump powers increased, EDFAs could move gracefully from enhancing single-wavelength systems to enabling WDM (wavelength division multiplexed systems) carrying many wavelength channels, which multiplied the effective value of each installed amplifier manifold times. In fact, EDFAs were expensive enough to help justify a migration to WDM. As reliable pump lasers at 980 nanometers were developed, EDFAs could move from reliance on pumping at 1480 nanometers, and achieve fundamentally better performance. As 1480-nanometer lasers were increased in power, longer wavelength sub-bands within the third window could be used. Finally, designers configured EDFAs to allow midstage access; that is, a preamp and postamp could be sandwiched around a lossy element without significant degradation in end-to-end system performance. This platform supported the use of dispersion-compensating modules, allowing higher data rates per channel (10 gigabits per second), and of optical add-drop multiplexers, which turned point-to-point systems into optical networks.

Ironically, these technology trends, which were supported by the commercial application of EDFAs, have led to the reinvigoration of a competing technology, Raman amplification. The threshold pump power requirements, which once appeared prohibitive, are now achievable because available laser pump powers have increased by ten times to meet the requirements of WDM systems with over 100 channels. In fact, Raman amplification may enable the next telecommunications revolution, ultra-longhaul (ULH) systems with transmitter-to-receiver distances of more than 1500 kilometers, even up to the 5000 kilometers of a coast-to-coast transmission in the U.S. To achieve this, engineers can realize distributed Raman amplification by injecting pump power into the fiber span to give gain (or loss reduction) along a large fraction of the span. They can thus maintain optical signal-to-noise ratios several times higher than in a pure EDFA system with the same span lengths, where the amplification occurs in a module at the end of the span. While most ULH proposals involve a mixture of EDFAs and Raman amps, at least one supplier has proposed an all-Raman solution using distributed Raman amplifi-

cation as well as discrete modules as direct substitutes for EDFAs.

The WDM backbone, carrying enormous amounts of data and voice traffic between more localized metropolitan and access communications networks, no longer represents a bottleneck in data transmission, and attention is shifting to the access network with the expectation that users will ultimately enjoy the advantages of bandwidth much greater than that at the time of writing. Business models for this market emphasize inexpensive technologies, and devices that cost little to manufacture and operate may have an overall advantage even if their technical attributes are less attractive than those of costlier alternatives. Thus SOAs are currently generating significant interest for their potential low-cost production and low operating power.

EDFAs, as the first practical implementation of optical amplification, drove the technological development of telecommunications in the last decade of the twentieth century, but their very success and influence may have sown seeds that will lead to their fall from prominence in the first decade of the twenty-first. Optical amplification is, however, definitely here to stay—without it, modern telecommunications systems would not function.

See also **Lasers in Optoelectronics; Optical Materials; Optoelectronics, Dense Wavelength Division Multiplexing; Telecommunications**

MARTIN HEMPSTEAD

Further Reading

Becker, P.C., Olsen, N.A. and Simpson, J.R. *Erbium-Doped Fiber Amplifiers*. Academic Press, New York, 1999.

Hecht, J. City of Light: *The Story of Fiber Optics*. Oxford University Press, Oxford, 1999.

Optical Fibers, *see* **Optical Materials**

Optical Materials

Optical materials were essential for many of the twentieth century's most significant technological accomplishments. Humans have been intrigued with the development of optical materials and light behavior for centuries. Thomas Alva Edison, Max Planck, Albert Einstein, Max Born, and Niels Bohr presented significant optical theories and innovations that provided a foundation for research, development, and application of twentieth century optical materials. Significant optical milestones included the introduction of lens coating in the 1930s and lasers in 1960; fiber optics emerged by mid-century, and thousands of scientists innovated and adapted uses for optical materials in the following decades. Engineers established the International Society for Optical Engineering (SPIE) in 1955 to coordinate professional efforts.

During the 1870s, German chemist Dr. Otto Schott sought to make glass with optical qualities, devising a lithium formula. By 1884, he produced optical glass specifically for lens and prism applications such as microscopes. With colleague Dr. Ernst Abbe and industrialist Carl Zeiss, Schott established the pioneering optical glass-manufacturing site at Jena, Germany. As a result, Germany was the main manufacturer of optical glass prior to World War I.

Initially, scientists were most concerned with the composition of optical glass. The production of optical glass and removal of impurities is crucial to achieve optimal performance. The homogeneity of optical glass differentiates it from most glass. Its consistent, predictable quality; strength and durability despite physical stresses and temperatures; precise images; minimization of distortion; and versatility for desired applications all make it ideal for optical usages. In the early twentieth century, German technologists advanced optical glass technology with such innovations as Schott's borosilicate and barium glass. By 1939, George W. Morey of the U.S. devised optical glass composed of tantalum, lanthanum, and thorium that increased refraction.

Before World War II, most optical glass manufacture resulted from clay pot melting processes. After the war, technologists focused on manufacturing aspects to improve optical glass. Some technicians used platinum pots and rare earth elements to create optical glass with minimal flaws (such as bubbles, weaknesses, and striae) where refraction differed. Japanese technologists at Hoya Corporation initiated a continuous melting technique in 1965 that produced flaw-free optical glass. Engineers worldwide devised better materials and melting processes, including electric melting, to create optical glass for electronics and cameras that were becoming lighter, smaller, and more precise because of optical glass lens advancements. Most modern optical glass is made from fused silica that is synthetically manufactured to enhance purity and quality.

In 1938, American Katherine J. Blodgett patented a layered soap film to produce transparent glass that did not reflect light rays and distort images in lenses because the soap and light waves

neutralized each other. Her findings inspired optical coating work to apply transparent thin films to lenses, glasses, and electronic devices. Scientists developed materials especially for those purposes. Most modern optical coating materials are metals, semiconductors, or insulators.

In the late twentieth century, plastic lenses replaced eyeglass lenses. In addition to being more shatterproof, plastic material such as polycarbonate enables technicians to apply or incorporate a wider variety of optical technology, primarily in coatings, to enhance vision. Each side of a plastic lens is layered with thin films that aid in making the lens invulnerable to water, glares, and scratches. A physical vapor deposition technique is used for the antireflective coating in which such metal compounds as aluminum oxide undergo vaporization in a vacuum chamber and several layers of atoms are deposited on a lens. A hydrophobic coating is layered on the lens by chemical vapor deposition or submersion in a solution. Engineers improved coatings that protect eyes from ultraviolet light.

The transparent vinyl thermoplastic polymer polymethylmethacrylate (PMMA) was first industrially produced in the mid-1930s after chemists initiated research the previous decade. PMMA is commonly referred to as acrylic, a word derived from the name acrylates, the polymer family that includes PMMA. This hard synthetic plastic cannot be shattered like glass and is used for various optical needs. In 1928, the company Rohm & Haas (based in Philadelphia and Darmstadt) began to sell acrylics based on resins made by Rudolph Fittig in the nineteenth century and adapted by German chemist Otto Rohm, for coating use as a celluloid substitute. At Imperial Chemical Industries, English chemists John Crawford and Rowland Hill made a harder acrylic in 1934 in a heating, pouring, baking, and cooling process to create acrylic sheets that were marketed as Perspex. Competition between acrylic manufacturers resulted in improved acrylic production methods and applications. Acrylic is more vulnerable to impact damage than polycarbonate in lenses, but PMMA is considered a valuable optical material for its superb clarity, electrical characteristics, and resistance to scratching. Engineers developed computer-guided looms to weave acrylic optical fibers. Acrylic is also used for LED and other optical displays. Lucite (DuPont) and Plexiglass (Rohm & Haas) are well-known trade names for acrylic optical products.

Because of its lightness, flexibility, affordability, and ability to be tinted with color, plastic is often shaped to make contact lenses. Leonardo da Vinci first suggested contact lenses in the early sixteenth century, and scientists developed similar glass lenses in the following centuries. In the mid-1930s, American Dr. William Feinbloom created hard plastic contact lenses. Californian Dr. Kevin Tuohy received a PMMA corneal contact lens patent in 1950. Industries such as Bausch and Lomb produced lenses, and researchers improved designs to resolve problems patients reported. English ophthalmologist Harold Ridley innovated the intraocular lens (IOL) in 1949 for lens replacement in cataract surgery. Although PMMA is used for IOLs, chemists developed softer acrylics and silicones that can be folded to insert in smaller eye incisions than required for PMMA IOLs.

Softer plastics gradually replaced PMMA as a lens material except for specific uses such as treating astigmatism. By the 1960s, contact lenses became more appealing when Czechoslovakian optometrists Otto Wicherle and Drahoslav Lim used the polymer hydroxyethylmethacrylate (HEMA) to make soft contact lenses. They selected materials that can absorb moisture to become flexible. By 1979, rigid gas-permeable (RGP) contact lenses containing primarily silicone (which permits oxygen to reach the eyes) were introduced. RGPs often correct vision problems that soft lenses cannot, last longer, and do not collect as much debris as soft lenses. Technological advancements in contact lenses included bifocals, disposable contacts, and improved material formulas that resulted in thinner lenses and increased oxygen access to the eyes.

Fiber optics enable telecommunications to span the globe swiftly and clearly. Glass fibers transmit infrared light pulses that are sounds or digital information transformed into light by semiconductor lasers. Optical fibers, made of a core inside a separate glass cladding, consist primarily of extremely pure silica. Scientists add fluorine and boron to the cladding and phosphorous and germanium to the core to manipulate those components' refractive index as needed for efficient light movement through fibers. The cladding prevents light from leaking out of the core by reflecting light within the boundaries of the core (total internal reflection).

Visionaries who proposed fiber optic technology in the 1920s and 1930s but did not follow through with their ideas included American Clarence Hansell, who received a patent for bundling glass fibers, and German Heinrich Lamm, who experimented with glass fibers for surgery. Narinder S. Kapany is often credited as fiber optics' inventor,

but his Massachusetts Institute of Technology doctoral advisor, Harold H. Hopkins, suggested Kapany's topic. Hopkins had been investigating glass fibers and wanted to bundle them to transmit images when he received a grant which he used to hire Kapany. The initial letter describing their work was published in *Nature* on 2 January 1954 beneath a letter about fiber bundling that Dutch scientist Abraham C.S. van Heel had submitted months before. Kapany wrote a *Scientific American* feature in 1960 and the first book on that topic.

Dr. Charles K. Kao is also cited as the innovator of fiber optics because of his research at the English ITT Standard Telecommunications Laboratories in the 1960s. Kao and Charles Hockham wrote about fiber optic possibilities. Kao was particularly interested in purifying optical materials such as silica compounds to remove metal impurities and improve transmissions. During the early 1970s, Corning Glass Works researchers Robert Maurer, Don Keck, and Peter Schultz developed a heat process to make extremely clear glass fibers from fused silica doped with germanium. Their achievement resulted in fibers with much lower attenuation (power loss with distance along the fiber), and fiber optic communications became feasible. Losses in optical fibers were much lower than in copper cables, and fewer repeaters meant lower cost systems. At the same time, researchers developed semiconductor lasers for use as fiber-optic light sources.

In addition to these pioneers, since the 1960s, many people and corporations, often fiercely competing to secure patents, have contributed to the advancement and distribution of fiber optics worldwide. Some researchers focused on multimode optic fibers instead of single-mode fibers, delaying the technology in some regions because of interference and noise caused by conflicting modes and waves that can disrupt transmissions. Toni Karbowiak's 1964 waveguide aided acceptance of optical fibers when they first became publicly available in the 1970s. On 19 May 1971, Queen Elizabeth II observed a fiber-optic video presentation, and other trials were held in the U.S. By 1975, a pioneering fiber-optic system was in use for communications by police in Great Britain. Engineers at British Telecom, Bell, and GTE introduced fiber optics into their systems.

Researchers advanced fiber optics quickly in the 1980s, improving services particularly for long-distance telephone calls. International service advanced in 1988 with the introduction of TAT-8, the initial transatlantic fiber-optic cable that provided increased circuits, greater capacity, and clearer signals than satellite and wire connections. More fiber-optic cables connected continents, with each fiber capable of transmitting several hundred million bits every second. During that decade, the University of Southampton's Dave Payne realized erbium would be the most useful amplifier material in fibers for clear, uninterrupted signals transmitted via ocean cables.

Fiber optics reduced the need for amplifiers due to attenuation. Unless water vapor is present, most silica in optic fibers does not absorb material sufficiently to interrupt signal movement. At higher wavelengths, silica is prone to absorb material, and substitute fiber sources such as fluorozirconate are often used. Engineers focus on solving dispersion problems related to silica interaction with frequencies when various parts of a signal move at different speeds in fibers toward receivers, as this often creates pulse interference.

Researchers are constantly advancing fiber-optic materials, systems, and applications to achieve greater speed, amount of information transmitted, and practical uses. For example, fiber optics enables surgeons to access and examine internal organs without performing surgery. Fiber optics help guide missiles, find earthquake victims, provide transportation signals, and act as sensors. Although many communities' telephones, Internet, and cable television services relied on fiber-optic networks, because such technology is expensive, few private homes had optic fibers by the end of the twentieth century and were connected to those systems by wires.

See also **Electronic Communications; Lasers in Optoelectronics; Materials and Industrial Processes; Optometry; Telephony, Long Distance**

ELIZABETH D. SCHAFER

Further Reading

Agrawal, G.P. *Fiber-Optic Communication Systems*. Wiley, New York, 1997.

Buck, J.A. *Fundamentals of Optical Fibers*. Wiley, New York, 1995.

Chaffee, C. D. *The Rewiring of America: The Fiber Optics Revolution*. Academic Press, Boston, 1999.

Fenichell, S. *Plastic: The Making of a Synthetic Century*. Harper Business, New York, 1996.

Hecht, J. *City of Light: The Story of Fiber Optics*. Oxford University Press, New York, 1999.

Izumitani, T.S. *Optical Glass*. American Institute of Physics, New York, 1986.

Kao, C.K. *Optical Fiber Systems: Technology, Design, and Applications*. McGraw-Hill, New York, 1982.

Kao, C.K. *Optical Fibre*. IEE Materials & Devices Series vol. 6, Peregrinus/Institution of Electrical Engineers, London, 1988.

Kao, C.K., Ed. *Optical Fiber Technology II*. Wiley/Institute of Electrical and Electronics Engineers, New York, 1981.
Kapany, N.S. *Fiber Optics: Principles and Applications*. Academic Press, New York, 1967.
Marker, AJ. III, Ed. Properties and characteristics of optical glass. *Proceedings of SPIE—The International Society for Optical Engineering*, 970, SPIE, Bellingham, 1988.
Meikle, J.L. *American Plastic: A Cultural History*. Rutgers University Press, New Brunswick, NJ, 1995.
Morey, G.W. *The Properties of Glass*. Reinhold, New York, 1938.
Mossman, S., Ed. *Early Plastics: Perspectives, 1850–1950*. Leicester University Press, London, 1997.
Mossman, S.T.I. and Morris, P.J.T., Eds. *The Development of Plastics*. Royal Society of Chemistry, Cambridge, 1994.
Park, D. *Fire Within the Eye: A Historical Essay on the Nature and Meaning of Light*. Princeton University Press, Princeton, 1997.
Seymour, R.B., Ed. *History of Polymer Science and Technology*. Dekker, New York, 1982.
Seymour, R.B. *Pioneers in Polymer Science*. Kluwer, Dordrecht, 1989.
Su, F., Ed. *Technology of Our Times: People and Innovation in Optics and Optoelectronics*. SPIE Optical Engineering Press, Bellingham, 1990.
Willey, R.R. *Practical Design and Production of Optical Thin Films*. Dekker, New York, 1996.
Yariv, A. *Optical Electronics*, 5th edn. Oxford University Press, New York, 1997.

Useful Websites

International Society for Optical Engineering: http://www.spie.org

Optoelectronics, Dense Wavelength Division Multiplexing

The cluster of technologies that are dense wavelength division multiplexing (DWDM) emerged during the last decade of the twentieth century. Multiplexing is the sending of separate signals with one transmitter in one optical signal, to increase data capacity. The pioneers of optical fiber communications had very early understood the potential of telecommunication fibers to carry more than one communication channel if the channel were on different optical wavelengths ("colors") but it took an army of researchers many years to realize the necessary technical solutions.

The advantages of DWDM are very clear—the technology offers a relatively straightforward path to exploiting the enormous inherent bandwidth of optical fiber. DWDM involves an array of modulated light sources at discrete wavelengths, which are combined ("multiplexed") onto a single transmission fiber. The signal may travel many hundreds of kilometers, being amplified every 80 to 100 kilometers by erbium-doped fiber amplifiers (EDFAs), which are continuations of the transparent path and require no separation of the channels. Ultimately, the channels are demultiplexed at the terminal equipment and each channel is individually detected and its modulated data converted to digital electronic format.

Until EDFAs emerged at the beginning of the 1990s, displacing electronic regenerators, DWDM was an expensive luxury. Electronic regeneration required the demultiplexing of channels at each node, so DWDM had little or no advantage over the use of a different fiber for each channel. The single channel-one fiber approach at least avoided expensive optical multiplexing, and allowed the use of any signal laser source, without tight wavelength control, for each channel, thus saving on the cost of maintaining a large inventory of expensive spare sources. The EDFA made it most economical to put all channels on the same fiber, saving cost at the nodes at the expense of terminal multiplexing and demultiplexing.

A number of technologies were required to make DWDM successful, and we will touch on each in turn.

- Since the channels are defined by the filter passbands in the optical multiplexers, light sources—semiconductor lasers—had to be stable in wavelength and available on a well-defined grid of wavelengths. They also had to be capable of modulation in a format that would be stable over the transparent links. In particular, they had to have acceptable "chirp" (phase distortion over the pulse) so that dispersion in the optical fiber would not degrade the pulse. The initial solutions were directly modulated diodes with integral grating-based distributed feedback (DFB) to lock the lasing wavelength.
- The multiplexers (which in reverse also served as demultiplexers) were of course key to the entire DWDM approach. Their function was to take a number of channels of different wavelengths each on a separate fiber and put them out on a single fiber. While this could be done with simple couplers such as fused biconic taper (FBT) couplers, the loss at each coupler is a factor of 2 in power, so high channel count DWDM would require very high laser powers. Since beams of light at different wavelengths can in principle be combined without loss, a more elegant and more scalable approach was to

use some kind of resonant coupling structure, for example, grating-assisted couplers or dichroic thin film filters.

- EDFAs suited for multichannel amplification had to be designed for high power, which required the development of more powerful pump lasers. The uniformity of gain across the wavelength band was also a challenge. While this was initially solved by careful positioning of the signal wavelengths in the flattest part of the erbium gain spectrum, the inevitable demand for more and more channels forced the use of gain-flattening filters in the amplifiers to tailor the gain spectrum. GFFs are optical filters with spectral loss curves engineered to match the erbium spectrum. Initially researchers manufactured them from carefully designed multilayer thin film filters, and later they used fiber Bragg grating (FBG) technology, when the grating community had developed techniques for fabricating complex chirped gratings.

The first DWDM systems were installed in submarine cables in 1990. Terrestrial systems followed later—they initially sported just four channels, each channel operating at 2.5 gigabits per second (Gbps), and were installed in about 1994. However, the expectations of long-distance telecommunications carriers quickly exceeded this unprecedented single fiber bandwidth of 10 Gbps, and an explosion of activity in the next few years pushed channel data rates quickly up to the next level—OC192 or 10 Gbps—and channel counts from 4 to 16 to 32 to 96, while the separation between adjacent channel wavelengths decreased accordingly. The capabilities of EDFA pump, multiplexer and GFF technologies had to be continuously upgraded to support this revolutionary increase in communication capacity. At OC192 rates, fiber dispersion became a problem and the new technology of dispersion compensation was also introduced, based on dispersion-compensating fibers that had been invented earlier in the decade. Dispersion compensation modules (DCMs) are typically sandwiched between EDFAs to compensate for their loss and thus minimize their impact on signal-to-noise ratio. Finally, the design of the fiber itself was modified to deliver better performance for the high-density high data-rate communications on the horizon. By the end of this explosion the potential capacity of new systems was about 1 terabit per second (10^{12} bits per second) per fiber, and the rediscovery of Raman amplification was set to extend system reach from 600 to 5000 kilometers.

With this breathtaking success under its belt, the telecommunications engineering community in 2000 confidently anticipated the continued rapid evolution of channel data rates to 40 Gbps, a doubling or more of the useable fiber bandwidth and the imminent fulfillment of the Raman promise. New challenging problems would have to be tackled—optical nonlinearities in the fiber (self-phase modulation (SPM), four-wave mixing (FWM), cross-phase modulation (XPM), etc.) and polarization mode dispersion (PMD) were the next hurdles to be overcome. New signal modulation formats and enhanced forward error-correction schemes would help to overcome these challenges. Unfortunately, the technological revolution had been accompanied by a massive expansion in fiber plant and high-performance system installation, fueled by the same expectation of continually exploding demand for bandwidth and solid revenue streams that had driven the investments in technology. While bandwidth continued to grow, albeit more slowly than predicted, the revenue did not materialize, and the fiber buildout turned out to have delivered a glut in capacity that devastated the telecommunications market as the bubble burst in the early years of the twenty-first century.

See also **Electronic Communications; Lasers in Optoelectronics; Optical Materials; Optoelectronics, Frequency Changing; Telephony, Digital**

MARTIN HEMPSTEAD

Further Reading

Laude, J.-P. *DWDM Fundamentals, Components, and Applications*. Artech House, 2002.

Ramaswami, R. and Sivarajan, K. *Optical Networks: A Practical Perspective*, 2nd edn. Morgan Kaufman, 2001.

Optoelectronics, Frequency Changing

Lasers produce monochromatic light; that is, light with a single frequency or wavelength. Many laser applications, such as atomic spectroscopy, depend on the ability of a particular laser to be frequency tunable. Fine tuning (small shifts in output frequency) can be achieved by adjustments of operating characteristics of dye and semiconductor lasers. In the area of fiber-optic communications, a large number of output wavelengths is desirable for wavelength division multiplexed (WDM) multichannel optical communication systems that utilize many wavelengths to increase data capacity. This

can be achieved by nonlinear effects that create new optical frequencies from fundamental frequencies (frequency doubling), or with tunable laser sources.

A nonlinear laser-induced effect, discovered by Peter Franken and co-workers in 1961, occurs when an intense laser beam propagates through a nonlinear optical medium (quartz, for example). Light at double the frequency of the input beam is produced, an effect known as frequency doubling or second harmonic generation. Second harmonic generation can usefully convert the coherent output of a fixed-frequency laser to a different spectral region. For example, the infrared radiation of a Nd:YAG laser operating at 1064 nanometers (infrared) can be converted into 532-nanometer visible radiation with a conversion efficiency of more than 50 percent. Novel laser sources produced by frequency doubling may offer advantages over existing laser sources that are bulky or inefficient; ultrashort laser pulses produced by frequency doubling or tripling may also have applications in the early twenty-first century for high-density optical storage, increased data transmission rates, and ultrafast spectroscopy of biochemical processes.

As the intensity of the incident light exceeds a certain threshold value, a nonlinear effect known as stimulated Brillouin scattering, which scatters light to different wavelengths, becomes important in fibers. Power is lost and the frequency-shifted wave may ''cross-talk'' with wavelengths in neighboring channels. Both of these effects degrade the optical signals and are undesirable.

Tunable dye lasers were discovered in 1966 by Peter Sorokin and John Lankard in the U.S. and Fritz Schäfer in Germany. Liquid lasers of this type can be made for almost any wavelength from the ultraviolet to the infrared, dependent upon the fluorescence of the dye. Dye lasers may be tuned, through 100-nanometer ranges or more, by either changing the concentration of the dye or by adding and turning a diffraction grating in place of one of the cavity mirrors.

The class of laser used in the majority of today's telecommunication systems is the semiconductor laser. This class of laser was first made from gallium arsenide (GaAs), but now they are commonly formed by a compound of elements from groups III and V of the periodic table (see Semiconductors, Compound). The end faces of the crystal (0.1 to 1 millimeter thickness) form mirrors that create the necessary cavity to trap the light and sustain stimulated emission.

Semiconductor lasers operate as either ''fixed'' (a single wavelength, or frequency) or ''tunable'' (offering coarse or fine tuning of many frequencies across a specific frequency band). The wavelengths of tunable semiconductor lasers depend upon both the properties of the III–V gain medium and the physical structure of the laser cavity surrounding the gain medium. Specifically, the length of the cavity (i.e., the distance between the mirrors) and the speed of light within the gain medium within the cavity determine a laser's wavelength. A semiconductor laser can therefore be tuned by mechanically adjusting the cavity length or by changing the refractive index (the speed) of the gain medium. Alternatively, light can be adjusted externally to the laser source, using micromachined elements such as micromirrors or actuators.

Tunable semiconductor lasers can be grouped into four types:

1. The ''edge emitting'' triad of distributed feedback (DFB)
2. Distributed Bragg reflector (DBR)
3. External cavity diode lasers (ECDL)
4. The ''surface emitting'' type known as vertical cavity surface-emitting lasers (VCSEL)

As one might suspect, edge-emitting devices emit light at the substrate edges, whereas VCSELs emit light at the surface of the laser diode chip.

Rather than placing the resonator mirrors at the edges of the device, the mirrors in a VCSEL are located on the top and bottom of the semiconductor material. Somewhat confusingly, these mirrors are typically DBR devices. This arrangement causes light to ''bounce'' vertically in a laser chip, so that the light emerges through the top of the device, rather than the edge. As a result, VCSELs produce beams of a more circular nature than their cousins and beams that do not diverge as rapidly. These characteristics enable a more efficient coupling of VCSELs to optical fibers. VCSELs, furthermore, benefit from single-process manufacturing and a relatively straightforward tuning process involving a microelectromechanical-systems (MEMs) cantilever arm placed directly above the optical cavity. Moving the arm a matter of a few micrometers up or down changes the frequency of the device by up to 5 percent.

The edge-emitter family features diffraction gratings etched on a single chip, as in DFB lasers; gratings placed near the active region of the laser cavity, in the case of DBRs; and one or two mirrors combined with a conventional laser chip to reflect light back into the cavity, as found in ECDLs.

DFB lasers can be tuned by controlling the temperature of the laser diode cavity. Because this technique requires large temperature differences, a single DFB has a small tuning range. However, it is possible to link multiple DFBs together to create multiple cavities and therefore wider tuning outputs. DBRs are actually variations of the DFB. In addition to having their grating (or mirror) section in a separate portion of the chip, a DBR has a gain section and a phase section. Tuning occurs when current is injected into the phase and mirror sections to change the carrier density of those two sections and, as a result, the wavelength of light they refract. As with DFBs, DBRs have a somewhat limited tuning range, but techniques have been developed to expand that capability; for example specialized gratings known as "sampled gratings" and grating and bidirectional coupler combinations. The ECDL achieves tunability by physically moving a wavelength selective element, such as a grating or prism, to tune the laser output. One method involves moving a reflector up and down relative to the surface of a diffraction grating, with the varying distances determining specific wavelengths. This particular tuning method gives it a wide tuning range, but a slow tuning speed compared to the other methods.

Because of their inherent design, specifically characteristics of the gain medium and cavity, VCSELs have lower power than the other three types and as a result are used principally for local or metro (metropolitan area) wavelength applications. ECDLs have the highest power output of the tunables discussed here and are used for long distance networks. The others lie somewhere in the "middle space" between VCSELs and ECDLs and are used for metro and regional applications.

As noted above, tunable lasers have emerged for use in optical communication systems and specifically "wave division multiplexing" (WDM) or dense WDM (DWDM) applications, which entail the process of sending many different wavelengths carrying information down a single fiber optic strand to increase data capacity. WDM emerged as a solution to the bandwidth crunch imposed on telecommunications by the ever-increasing growth of the Internet and other broadband applications and it presents an exciting and new enabling application for tunable lasers in local, regional and long-distance networks.

In future all-optical networks (i.e., those without conversions to electronic signals), fiber optic systems will require tunable lasers that can provide a signal into any WDM channel and that can switch among channels, tuning to a new output wavelength in nanoseconds. In 2001 and 2002, some of the first true tunable semiconductor lasers moved from prototypes to early production.

See also **Lasers in Optoelectronics; Optoelectronics, Dense Wavelength Division Multiplexing**

CHARLES DUVALL

Further Reading

Bruce, E. Tunable lasers, *IEEE Spectrum,* 35–39, 2002.

Chang-Hasnain, C.J. Tunable VCSEL. *IEEE Journal on Selected Topics in Quantum Electronics*, 6, 6, 978–987, 2000.

Schäfer, F. P., Schmidt, W. and Volze, J. Organic dye solution laser, *Appl. Phys. Lett*. 9, 306–309, 1966.

Sorokin, P. P. and Lankard, J.R. Stimulated emission observed from an organic dye, choloalurinum phthalocyanine. *IBM J. Res. Develop*., 10, 162–163, 1966.

Optometry

The word, "optometry" was introduced to ophthalmology in 1904 and is important as an adjunct to that medical specialty because it focuses on examination of the eyes, analysis of vision, and the prescription of corrective or preventive measures for any deficits or problems. An optometrist will refer a patient to an ophthalmologist when pathologies of the eye are found during examination of visual function. Most of the technology used in optometry involves measurement and lenses. The only pharmaceuticals used are pupil dilators, drugs used to make the pupils larger so that the posterior parts of the eye can be examined.

Optometry developed from a split in the optical profession that took place in the nineteenth century that resulted two kinds of opticians, refracting and dispensing. Refracting opticians became optometrists. In Canada and the U.S., an optometrist must obtain an undergraduate degree and then undertake four years of specialized graduate education plus clinical and resident training to become an OD, or doctor of optometry. By contrast, a dispensing optician's training is from six months to two years (which varies by location), and does not necessarily require formal education. The optician can apprentice with an ophthalmologist, optometrist, or another optician. An optician makes or dispenses lenses and eyeglasses; an optometrist performs eye examinations.

Technologies of optometry can be subdivided into two categories: development of eyeglasses, contact lenses, and coatings; and instruments of measurement.

The idea that a convex lens could be used to magnify was known to the ancients. The use of a lens as a magnifying glass to aid vision is attributed to Roger Bacon in Opus Majus (major work), written in 1268. Soon after, Italian monks D'Armate (1284), da Rivalto (1306), or possibly a competitive contemporary, da Spina, invented eyeglasses. This early device was hinged and held two magnifying spheres that rested uncomfortably on the nose without support.

In 1303, Bernard of Gordon (Montpellier, France) fashioned a pair of spectacles with a fixed bar, but it took 400 years of various experiments, from ribbons looped around the ears, to weights that balanced glasses on the nose, until a design was engineered to allow the glasses enough stability to free the reader's hands. Corrective lenses were developed in the early seventeenth century.

The popularity of spectacles grew throughout Europe, but because they were promoted by itinerant peddlers with little or no education, medical professionals were hesitant to endorse such products. Ironically, the growth of optometry was not because of oculists or physicians but in spite of them. Each time these unlicensed vendors and peddlers were able to fit spectacles to someone visually impaired, it increased their knowledge and skill.

By the nineteenth century, telescopes, microscopes, and cameras, and optical instruments required lenses for their use. However it was the adaptation of these lenses to that gave optometry its refined technology. An early ophthalmometer that used a small telescope was developed by E. Javal and H. Schiötz in Germany in 1881 and later modified into an astigmometer. The ophthalmoscope was invented by Hermann Von Helmholtz (Germany) in 1851; the retinoscope was introduced by Cuignet in 1873 in France; and Placido's keratoscope appeared in 1882. By the beginning of the twentieth century these instruments were in place to fit the established practices of optometrists.

The greatest changes to optometry in the twentieth century were in modifications of the instruments for diagnosis and the proliferation of materials that rendered lenses—for both glasses and instruments of measurement—thinner, lighter, more durable, safer, and more accurate. The four major instruments used by optometrists are the retinoscope, or skiascope, the slit lamp, the phoropter (introduced in 1938), and the ophthalmoscope.

The retinoscope illuminates the retina. It is used to test astigmatism, farsightedness, and nearsightedness. Early machines used a light source directed from a mirror. In the early 1900s, the technique was refined and termed spot retinoscopy. The operator recorded the direction of movement of the light on the retinal surface and the angle of the light rays that emerged as a reflex from the patient's eye, a phenomenon known as refraction. In 1926, ophthalmologist Jack Copeland inadvertently dropped the spot retinoscope he was using. In what might be considered a serendipitous event, he recognized that the damaged equipment could still be used and that it produced a clearer image of the patient's eye. Thus the streak retinoscope was invented. He patented it in 1927, and it remained unchanged until 1968 when a cordless retinoscope was developed by Optec (the Optec 360). In the twentieth century, independent light sources—an electric bulb or battery-operated machine—were employed instead of mirrors. In 1992, a new retinoscope was developed which did not require measurement by the operator at all. Instead, the patient's eye was measured against computerized predetermined calibrations and then transposed into meaningful data that was used by the optometrist to determine the patient's prescription if correction were needed and the lenses that would be required.

The slit lamp looks like a very elaborate microscope. It sits on a table and has a chin and headrest attached. The light source with controls is on a movable arm and the operator is able to adjust brightness, the size of the beam of light, and insert different filters. It uses a set of hand-held lights directed at the front of the patient's eye and examines the eyelid, the sclera (white of the eye), conjunctiva (mucous membranes), iris (colored part of the eye), lens, and the cornea (outer covering). It is called a slit lamp because its light source shines as a slit. This specialized magnifying microscope-type device was invented by Allvar Gullstrand in 1911 in Sweden. Later, attachments were added to include a camera for taking photographs, a tonometer for measurement of intraocular pressure (significant in glaucoma), a pachymeter to measure corneal thickness, and a laser treatment module for the ophthalmologist. It also had changeable lenses.

The phoropter is imposing in appearance. It is suspended from a bar and looks like a giant pair of glasses with five lenses on each side. The machine is used to test vision by moving sets of lenses in front of the patient's eye. Traditional phoropters are used to both measure and correct vision in order to derive an optical prescription for the patient. In the last decade of the twentieth century, a MEMS

(microelectromechanical semiconductor)-based adaptive optics phoropter (MAOP) was developed that used adaptive-optics technologies and a deformable mirror. Interestingly, this technology was originally developed for astronomy applications. Of benefit to the optometrist is that the newer machine requires less space and automatically calculates the numbers required for the vision correction prescription.

The von Helmholtz ophthalmoscope consisted of a handle and a group of lenses that could be interchanged. It is an instrument that allows the doctor to look inside a person's eye and view the optic disc. Modern ophthalmoscopes are either direct or indirect. The direct is a hand-held instrument with a battery-powered light source. It has a series of lenses that are dialed in to focus the doctor's view of the central retina. The indirect ophthalmoscope is used to examine the entire retina. This instrument is worn on the doctor's head and another lens is placed in front of the patient's eye. The ophthalmoscope is used by both optometrists and ophthalmologists.

Materials

Eyeglasses were made exclusively from glass until the 1950s when the revolution in plastics began. Two companies pioneered the development of a thinner, lighter, and flatter lens made from plastic: American Optical (AO) and Columbia Southern Chemical Company. In 1937, AO produced a polymethyl methacrylate (PMMA), a hard plastic material that could be molded for lenses, but it was not scratch resistant and often distorted under high temperatures. During World War II, Columbia Southern Chemical Company produced a series of 200 polymers. From these, number 39, the chemical composition of which is allyl diglycol carbonate (ADC), had ideal qualities for a lens. It did not soften or distort at high temperatures and was scratch resistant. It was cast rather than molded. Named CR-39, this became the industry standard for eyeglass and contact lenses.

In 1971, a flexible plastic contact lens containing a gel was developed that allowed the wearer to change the lens. The advantage was cosmetic and economical. Eye color could be changed with the lens and the price was greatly reduced, thus concerns about loss or breakage were less of an issue. In 1978, a rigid gas permeable lens (RGP) was developed that could be custom fitted to the cornea. By 1983, they were available for commercial distribution. Although they were hard, they allowed for the exchange of oxygen on the surface of the eye and were particularly advantageous to people who had undergone surgery for cataract removal. An extended-wear plastic, developed in 1981 allowed the wearer to keep the lens on the eye for longer than 24 hours and even sleep with it. These could remain as long as one week without change. A few years later, the disposable contact lens was on the market. By the late 1990s, the rigid gas-permeable contact lenses formula of fluorosilicone acrylate was modified. At the turn of the century, contact lens possibilities included disposable tinted lenses that could be used for two weeks; disposable lenses with ultraviolet ray protection; and multifocal disposable soft lenses that made a contact "bifocal" possible, thereby moving from eyeglasses and contacts to contacts alone.

See also **Ophthalmology; Optical Materials**

ELIZABETH D. SCHAFER

Further Reading

Garrison, F.H. *An Introduction to the History of Medicine*, 4th edn. Saunders, Philadelphia, 1929.

Gregg, J. *The Story of Optometry*. Ronald Press, New York, 1965.

Lyons and Petrucelli. *Medicine, an Illustrated History*. Abrams, New York, 1987.

Stein, H., Slatt, B. and Stein, R. *The Ophthalmic Assistant: A Guide for Ophthalmic Personnel*. Mosby, 2000.

Useful Websites/Further Information

http://web.grinnell.edu/groups/sca/glasses/glasses.htm
http://www.contactlens.co.uk/education/public/history.htm
http://www.contactlenses.org/rgpvsoft.htm
http://www.mse.utah.edu/students/MSE5471/ContGlas/Glasses%20Processing.htm
http://spectacle.berkeley.edu/class/opt10/lect2.shtm

Many institutions and companies can be consulted for information on specific machines. A few are: The Lawrence Livermore National Laboratory in California; Bausch & Lomb in Rochester, NY; Boston Micromachines Corp., Watertown, MA; Sandia National Laboratories, Livermore, CA; and Wavefront Sciences, Albuquerque, NM.

Organ Transplantation

Organ transplantation is a specialized case of the transplantation of living body tissue. In the 1880s scientists and surgeons started developing a concept of organ replacement that formed the basis of the technique. According to this concept it is possible to treat complex internal diseases by replacing the lost function of a particular organ. The roots of organ transplantation lay in the ever more sophisticated surgical strategy of removing

diseased tissues. Surgeons had noticed that the lack of particular organ tissue led to the development of specific disease phenomena, but reinserting the tissue into the body reversed disease development.

For the general acceptance of this concept it was significant that surgeons and physiologists were able to create and stop disease symptoms at will, using experimental animals under the controlled conditions of the laboratory. The first organ to be examined in this way was the thyroid gland. It was also the subject of the very first organ transplant by the Swiss surgeon and Nobel laureate Theodor Kocher in 1883. The principle was then applied to other organs, starting with other endocrine glands—pancreas, testes, ovaries, and adrenal glands. In 1905 Alexis Carrel and C.C. Guthrie in New York City carried out the first heart transplant in a dog. In 1906 Mathieu Jaboulay in Lyons performed the first kidney transplant in a human being, using an animal kidney.

The technical aspects of organ transplantation were mastered in the first decade of the twentieth century when Carrel developed a reliable and effective technique for suturing blood vessels. Carrel was awarded the Nobel Prize in 1912 for his work on vascular surgery and organ transplantation. With growing technical perfection it became clear that organ transplants between different individuals (allotransplantation) normally failed because of some specific factor associated with the biological identity of individuals. Some researchers ascribed the rejection of foreign tissue to the same mechanism that was also responsible for the body's defense against infectious agents. In the 1920s the German surgeon Georg Schöne introduced the notion of transplant immunity to describe the phenomenon. However, all attempts to prevent the rejection of allotransplants by suppressing the immune reaction or by selecting suitable donors failed. As a result of the inability to overcome these difficulties, organ transplantation was gradually abandoned in the course of the 1920s.

In 1945 a new phase in the history of organ transplants was initiated when surgeons at the Peter Bent Brigham Hospital in Boston transplanted a kidney from a dead donor to a woman suffering from renal failure. Even though this and subsequent transplants failed, the American surgeons did not abandon their efforts. In 1954, again in Boston, a patient with renal failure was given a kidney from his identical twin brother. The transplantation was successful, and in 1990 the operating surgeon Joseph E. Murray was awarded the Nobel Prize.

In the latter case it was possible to avoid transplant rejection by the selection of an appropriate donor. In order to make allotransplanation applicable on a broader scale, however, surgeons had to pursue a different strategy, which consisted in suppressing the recipient's immunological reaction against the transplant. In the 1940s, Peter Medawar and Macfarlane Burnet had described the principles of immunological rejection, following wartime work on tissue grafting. Initial trials using x-ray radiation for immunological suppression proved too damaging for the recipient, so further attempts concentrated on chemical immune suppression. In 1962 the first successful kidney transplantation from a nonrelated donor was performed in Boston. Immune suppression had been achieved by the antimetabolic agent azathioprine. This approach was subsequently perfected, one of its milestones being the introduction of the immune suppressor cyclosporine in 1982, which enabled more effective but simultaneously more selective suppression of tissue rejection. At the same time, efforts to select suitable organs from nonrelated donors were being made with the help of tissue typing using the human leucocyte antigen (HLA) system as a marker of compatibility. The allocation of transplants was organized by special organizations such as Eurotransplant, founded in 1967 to enable distribution in Austria, Belgium, Luxemburg and Germany.

Apart from kidneys, other organs were also soon transplanted in humans. Most spectacular was the first successful heart transplant in a human performed by Christiaan Barnard in Cape Town, South Africa in 1967. Because of technical and biological difficulties, however, heart transplantation was almost stopped during the 1970s, only to be resumed after the introduction of cyclosporine in the 1980s. At that time the transplantation of other organs, such as the liver, pancreas, and lungs were becoming more successful and therefore a more popular therapeutic option.

However, despite the development of new drugs for immune suppression and the introduction of immunological means to influence rejection (e.g., antilymphocyte globulines), the maintenance of long-term function of transplanted tissue is still considered an unresolved issue in the field. The other main problem concerns the procurement of organs. In the late 1960s, organ procurement from dead donors was regulated by formalizing criteria for the diagnosis of brain death. Brain death is a state in which the brain has died, but the rest of the body is kept alive with intensive care measures

such as artificial ventilation. Regulations were issued in different countries. For instance in the U.S. the so-called Ad Hoc Committee of the Harvard Medical School published a set of guidelines in 1968. Nonetheless, donation rates in no way kept up with the demand for organ tissue. Living donor transplants, which would be a viable alternative, are largely restricted to the kidneys, though split liver transplants have been performed with hepatic tissue from living donors. Another strategy to relieve organ shortage is the use of animal tissue. The procedure is called xenotransplantation and was tried from the very beginning of transplant medicine in the late nineteenth century. Results, however, were poor, and despite a number of research and development programs with pig organs in the 1980s and 1990s, xenotransplants do not seem to be a realistic option for the near future. The idea of growing tissues or even whole organs, which had been pursued by Alexis Carrel together with the engineer and aviator Charles Lindbergh in the 1930s, became popular again in the 1990s and led to the development of the field of tissue engineering. However, even if surgery should overcome all technical obstacles, transplantation will continue to raise a number of relevant cultural and bioethical issues concerning personal identity and the definition of human life. Organ transplantation is, after all, a technology that transcends boundaries of the individual body that had previously been taken for granted.

See also **Blood Transfusion and Blood Products; Immunological Technology**

THOMAS SCHLICH

Further Reading

Brent, L. *A History of Transplantation Immunology*. Academic Press, San Diego, 1997.
Fox, R.C. and Swazey, J.P. *The Courage to Fail. A Social View of Organ Transplants and Dialysis*. University of Chicago Press, Chicago, 1974.
Fox, R.C. and Swazey, J.P. *Spare Parts. Organ Replacement in American Society*. Oxford University Press, New York, 1992.
Hakim, N.S. and Papalois, V.E., Eds. *History of Organ and Cell Transplantation*. World Scientific, 2003.
Küss, R. and Bourget, P. *An Illustrated History of Organ Transplantation. The Great Adventure of the Century*. Laboratoires Sandoz, Rueil-Malmaison, 1992.
Lock, M. *Twice Dead. Organ Transplants and the Reinvention of Death*. University of California Press, Berkeley, 2002.
Moore, F.D. Give and Take: the Development of Tissue Transplantation. W.B. Saunders, Philadelphia, 1964.
Schlich, T. *Die Erfindung der Organtransplantation. Erfolg und Scheitern des chirurgischen Organersatzes (1880–1930)* [The Invention of Organ Transplantation: Success and Failure of Surgical Organ Replacement]. Campus, Frankfurt, 1998.
Woodruff, M.F.A. *The Transplantation of Tissues and Organs*. Charles C. Thomas, Springfield, 1960.

Organization of Technology and Science

The composition and configuration of institutions supporting technological advance changed dramatically over the course of the twentieth century. While national patterns differed substantially in the organization of science and technology, most industrialized nations experienced a few common trends. In the first half of the century, corporate research laboratories supplanted the workshops of independent inventor–entrepreneurs as the focal point of innovative activity. The establishment of corporate research laboratories and the increasing employment of engineers and scientists in industry coincided with, and to some extent facilitated, the expansion of institutions contributing to scientific and technological advance, including universities, government bureaus, and private research institutes. Following the end of World War II in 1945, institutions supporting science and technology expanded and new patterns of funding and coordination emerged. National governments, even those without significant military research and development (R&D) programs, exercised greater influence over the network of research institutions within their borders. Large-scale projects and collaborations across disciplines and institutions became common. In the third phase, beginning in the 1970s, industry reduced its support for basic research, shifted technical resources to divisional research programs, and increasingly entered into national and international strategic alliances. The influence of governments over the content of research in private industry diminished somewhat. Corporations, however, began to exert greater influence over research in universities. Market mechanisms influenced the organization of technology and science at the end of the twentieth century to a greater extent than during the previous period.

Differences in funding patterns, institutional roles and responsibilities, and relationships among institutions have accounted for national variations in the organization of technology and science. Questions relevant to discerning national variations include:

- Where is the locus of innovative activity?
- To what extent and through what mechanisms have markets, interest groups, and gov-

ernment agencies influenced the content and direction of private research activity?

- What distinctions exist in the motivations and goals of research in universities, firms, and government laboratories?
- What is the character of relationships between and among universities, firms, and government agencies?

While a systematic categorization of nations according to their organizational patterns is beyond the scope of this essay, these questions provide a basis for limited comparisons on discreet issues discussed here. In this entry, major changes in the organization of technology and science during the twentieth century will be explored, with sensitivity to national variations.

The Emergence of Institutions Supporting Technological Advance

Formal institutions supporting technological advance were in their infancy at the start of the twentieth century. In the late nineteenth century in the U.S., new technologies emerged primarily from the activities of independent inventors, such as Thomas Edison, and mechanics working in machine shops. Testing laboratories, such as the Pennsylvania Railroads, supported research to standardize equipment and improve materials, and universities remained almost completely disconnected from the process of technological change. The situation in Germany was different. The German chemical and electrical industries first began to exploit scientific knowledge systematically for technological purposes around 1880. German companies had the advantage of a national system of universities and state-sponsored research institutes with unparalleled capabilities in scientific education and research. Large firms, such as Bayer and Hoechst, funded individual professors in universities who, along with their graduate students, conducted studies relevant to the concerns of their sponsors. Most importantly, German firms hired graduate students with doctorates in chemistry and physics and established the first science-based industrial research laboratories. The German research model thus consisted of science-based research as a distinct corporate function and a network of scientific institutions that provided corporations access to the latest advances in scientific knowledge and a steady supply of new recruits for their industrial laboratories.

Leading corporations in the U.S., including General Electric, American Telephone and Telegraph, DuPont, and Eastman Kodak, established science-based research as a distinct organizational function in the first two decades of the twentieth century. Competitive threats from national and international rivals and fear of government antitrust action were among the reasons these companies established centralized research laboratories. To recruit and retain talented scientists, many of whom were trained in European universities, the managers of these pioneering laboratories often found it beneficial to provide their researchers with a modicum of freedom to choose their own research topics and to publish the results of their work. This tension between professional commitments and the goals of the corporation remained a major issue within industrial laboratories throughout the twentieth century. Nonetheless, the physicists and chemists that staffed the early laboratories generally conducted basic research related to their respective companies' core technologies.

While some scientists in American universities looked askance at their colleagues and students who sought employment in industry, applied science and academic engineering programs developed explicitly to meet the technical needs of industry. The chemical and chemical engineering departments at the Massachusetts Institute of Technology (MIT) grew large during the 1920s on consulting fees and corporate-sponsored research. Engineering programs at state universities such as Purdue University, the University of Illinois, and the University of Michigan, actively engaged in research relevant to the needs of industry. The chemical, electrical, and aeronautical engineering programs of such schools trained the legions of engineers who populated the production, development, testing, and even the research organizations of industrial corporations. Furthermore, the role of American universities in training scientists who found employment in corporate laboratories increased steadily throughout the first half of the twentieth century, and the pattern of collaboration between university scientists and industry spread to other fields such as the biomedical sciences.

In the American context, therefore, a large and diverse network of universities with strong training and research capabilities served as a foundation for the growth of industrial research. The British educational system, by contrast, did little to promote the expansion of research in industry. Limited government financial support for engineering education meant that Britain produced far fewer university-trained engineers than Germany and the U.S. Furthermore, university scientists

lacked incentives to develop relationships with corporations, and the linkages between the two remained tenuous throughout the first half of the century. Instead of developing the engineering and applied research capabilities of universities, the British government established cooperative research associations, which conducted routine studies to improve processes, but developed few new technologies. Under the British government's Department of Scientific and Industrial Research, the associations proved to be poor substitutes for corporate research. While some companies in cutting-edge industries developed in-house research capabilities, the focus of government policy on the promotion of cooperative associations and the isolation of science in the British context prevented the development of strong links between education, research, and inventive activity.

A handful of government agencies in the U.S. provided technical services similar to those of the British cooperative research associations. In 1901, Congress created the most prominent of these agencies, the Bureau of Standards, to establish, test, and maintain standards for industry, government, educational institutions, and scientific interests. The Bureau maintained programs in a wide range of fields including electricity, metallurgy, weights and measures, chemistry, and instruments. It generally limited its research program to solving technical problems related to the development of standards, but in dispersed industries that lacked research capabilities, such as the building trades, the Bureau supported research that supported technological advance more broadly. The Bureau of Mines conducted research to improve safety conditions and to improve efficiency and reduce waste in the mining and petroleum industries. The wind tunnels and laboratories of the National Advisory Committee for Aeronautics (NACA) served as centers of research and testing for the budding aviation industry. Although these organizations provided crucial services to industry in the years prior to World War II, their influence on the pace of technological change and the relations among research institutions was modest.

Other institutions that contributed to the advance of science and technology included independent research institutes and private foundations. Institutes such as Arthur D. Little, the Mellon Institute, and the Battelle Institute provided technical assistance to private corporations on a contract basis. Although the institutes were created in many instances to serve small firms, large firms turned to them for process research and help with routine technical problems, such as analyses of the properties of metals and chemicals. Private foundations operated independently of industry and supported basic research and education. Built on the wealth of prominent industrialists, these organizations possessed the means to set research agendas and to influence the national research environment. The Carnegie Institute of Washington, for example, supported individual and collaborative studies in the fields of genetics and terrestrial magnetism. The Rockefeller Foundation of New York supported basic research in the biomedical sciences and worked to spread medical cures and public health measures throughout the world. Following World War I, the Carnegie Corporation established a large endowment to support the National Academy of Sciences, and the Rockefeller Foundation began funding postgraduate education in physics and chemistry through the National Research Council.

Prior to World War II, support for science isolated from industrial concerns grew steadily in the U.S. but remained small in comparison with Europe. The lack of an institute for basic scientific research supported with government funds comparable to the Royal Institution of London or the French Academy of Sciences troubled the leaders of the scientific community who sought to raise the esteem of their profession. Failed efforts at achieving such ends, including the creation of the National Research Council during World War I and the Scientific Advisory Board during the Great Depression, did little to dampen the scientific community's quest to increase federal support for basic research. Whereas professional engineers in the U.S. had found accommodation with industry by the early twentieth century, scientists, even those working for and within industrial laboratories, remained ambivalent about their relationship with business. Despite such ambivalence, the American institutional context allowed for the institutionalization of science in industry and the formation of alliances between universities and industrial laboratories.

World War II and the Cold War

World War II represented a watershed in the organization of technology and science. The unprecedented technological achievements of the wartime mobilization, including the proximity fuse, radar, and the atom bomb, erased all doubts about the potential technological benefits of increased collaboration between industry, the uni-

versities, and government. Following the war, Vannevar Bush's famous report to the president, entitled *Science, the Endless Frontier* provided a formal justification for the establishment of a National Science Foundation in the U.S. and for increased federal funding of basic research in universities. Throughout much of the world, national military establishments came to occupy an important role in shaping the national scientific and technological infrastructure. Military demand for new weapons and government funding for science drove the expansion of existing institutions, facilitated the development of new methods for organizing research projects, and inspired the creation of new institutions for promoting, funding, managing, and conducting research.

Massive government expenditures for R&D and procurement provided the context for the expansion of corporate research following the war. A great many corporations hired teams of scientists to pursue basic scientific research and established new laboratories separate from engineering and production operations. Underlying this trend was a general misunderstanding, perpetuated by Vannevar Bush, of the innovation process. At the core of Bush's "linear model" of research was the notion that scientific research alone provided the basis for the development of new technologies. Innovation, according to the model, proceeds in a stepwise fashion, from research to development to production and then marketing. Unimpeded by market concerns and potential applications, scientists left on their own would produce fundamental ideas that applied scientists would then effortlessly transform into new technologies. Corporations that adhered to this model experienced great difficulty producing innovations, and many abandoned their research laboratories altogether by the early 1970s.

Although it had earlier precedents, "big science" emerged as a major theme in the organization of technology and science following World War II. Many activities with distinct organizational features have been lumped together under the term "big science," and the term is often used loosely to refer to any scientific or technological endeavor that involves vast financial resources and large numbers of researchers. Geographically concentrated projects centered on massive instruments and dedicated to advancing scientific knowledge, in this view, fall into the same category as research programs involving multiple laboratories dispersed over great distances, or even projects with a distinctly technological objective. In considering the organizational characteristics of large-scale science and technology, it is useful to distinguish such projects by their objectives (e.g., scientific or technological advance), geographic characteristics (e.g., concentrated or dispersed), and managerial structure (e.g., hierarchical or diffuse). Such categories, to be sure, may not fully capture the complexity of some projects, which, for example, may be centered in a single location but involve networks of individuals that stretch across continents. Nevertheless, distinguishing such projects by these dimensions allows a more complete sense of the diversity of approaches to the organization of large-scale technological and scientific projects in the postwar period.

The Manhattan Project is widely regarded as the wartime exemplar of big science. Under the leadership of General Leslie Groves, the U.S. Army's Manhattan Engineering District assembled and coordinated the activities of academics scientists, industrial firms, and construction contractors for the development of the atomic bomb. Even though a wide range of academic and industrial institutions spread throughout the country contributed to the project, Groves maintained tight control by compartmentalizing different aspects of the project. Groves ultimately had authority to control the distribution of funds and to transfer tasks to other groups if he did not get the results he wanted when he wanted them. Therefore, despite the role of certain sites as central nodes in the network of institutions contributing to the project, the sense of teamwork and community that evolved in certain compartments, and the scientific content of some of the work, the project was geographically dispersed, hierarchically controlled, and technologically oriented.

Big science projects in the postwar era, especially in the fields of astronomy and physics, were often organized around large instruments and directed toward scientific goals. In the case of nuclear physics, scientific entrepreneurs located in universities and government laboratories in the U.S. harnessed the resources of the state to fund the creation of increasingly larger particle accelerators. The construction and maintenance of accelerators, at places like Stanford, Berkeley, and Brookhaven, required advanced engineering knowledge and close collaboration between theoreticians, experimentalists, and engineers. Some accelerators were available as a shared resource for use by multiple research groups—a pattern also common in the field of astronomy. Particle physics research in Europe differed from the U.S. in a number of ways. Most important was the concentration of resources at a single institution, the

European Center for Nuclear Research (CERN). The multinational effort avoided military influence experienced in some U.S. laboratories, but it also experienced greater operational setbacks due to a much more rigid separation between scientists and engineers. During the 1950s, the Joint Institute of Nuclear Research in Dubna, Russia, served as the center of research for physicists in communist countries. As in the U.S., the Russians maintained particle accelerators at a number of locations and used them for military-related research.

The instruments produced to answer scientific questions or to coordinate scientific activities sometimes found applications beyond the laboratory. Nuclear magnetic resonance imaging, which physicists developed to measure the movement of nuclei, has been used in analytical chemistry and diagnosis of medical patients. The design of ARPAnet reflected the decentralized structure of the community of computer scientists and engineers that created and used it. In a few decades, the open-ended computer network moved from a tool for sharing data among researchers to a platform for economic transactions, personal communication, and entertainment, called the Internet.

National endeavors for the creation of large-scale technologies, such as Project Apollo, more closely resembled the Manhattan Project in organization and orientation than traditional big science projects. In the case of Project Apollo, the civilian administrators of the National Aeronautics and Space Administration (NASA) had responsibility for mobilizing and coordinating a geographically dispersed network of contractors, advisory committees, university researchers, in-house research centers, and government bureaucracies. A well-defined mandate backed by tax dollars and political support gave NASA administrators the authority they needed to protect researchers from external disruptions to prevent competition from within their network of bureaucracies and contractors from undermining the project. The Soviet space program, like the American program, was characterized by both centralization of resources and authority and high levels of competition among different interests within the space bureaucracy. Proposals for new spacecraft usually emerged from design bureaus that competed against one another in a formal review process. A scientific-technical council reviewed the proposals and made recommendations to a military–industrial commission for the final decision. Party politics played a role in the process. Subsumed under the Soviet military bureaucracy, overwhelmed by competition among design bureaus for limited resources, and lacking the narrow focus of the U.S. space program in the 1960s, the Soviet program lost the race to the moon.

Managers of large-scale development projects developed new techniques to cope with the organizational complexities they faced. From the U.S. Air Force's Atlas missile program emerged a number of concepts, including concurrency, which involved the pursuit of research, design, and production engineering in parallel on large-scale projects. Concurrency required the development of systems analysis capabilities in order to specify technical details at the start of a project. Concurrency and systems analysis were among the techniques incorporated into U.S. Air Force regulations in the early 1960s under the name configuration management. With configuration management, the U.S. Air Force required contractors to share cost estimates and designs with program managers and document changes that occurred during the course of a project, all of which were subject to approval by a configuration control board. The centralized, hierarchical reporting system imposed additional layers of bureaucracy on projects, although some observers claim that it controlled costs and forced contractors to abide by schedules and performance standards.

Collaboration, Internationalization, and University–Industry Relations

The decision of the U.S. Congress in 1993 to terminate the Superconducting Supercollider Project, a particle accelerator of unprecedented power, served as evidence of physicists' declining influence on science policy, but it did not mark the end of big science. Even though it relied on networks and multiple research sites rather than hierarchical communities amassed around large-scale instruments, the biogenetics community's Human Genome Project represented a continuation of the big science tradition. A short-lived defense conversion program following the fall of the Soviet Union in 1991 did little to halt the long-term growth of military budgets. With high levels of spending on research and costly weapons projects such as the Strategic Defense Initiative and a nanotechnology initiative, the U.S. Department of Defense assured its continued influence over the national research agenda. However, new patterns in the organization of science and technology began to emerge in the final decades of the twentieth century.

The failure of isolated research laboratories to make good on their promises to transform scien-

tific ideas into marketable technologies, as well as a general decline in funding for military research and basic science in the 1970s, laid the foundation for new patterns of research to emerge. Disillusioned companies either eliminated their centralized research laboratories or reorganized their research programs by creating closer links between research and product development and engineering and shifting research to corporate divisions. By the 1980s, decentralization and collaboration with external partners replaced isolated science as the central organizing principles of industrial research. International competitive threats served as the justification for the creation of industry-wide consortia. With significant financial backing from the military, the U.S. semiconductors industry established the Semiconductor Manufacturing and Technology Institute (SEMATECH) to improve commercial semiconductor manufacturing capabilities. Research consortia proliferated throughout the 1980s and 1990s. Great Britain established its LINK program in 1997 to support the establishment of consortia in precommercial research across the spectrum of advanced technologies. The European Union's Brite–Eura program supported consortia dedicated to the improvement of manufacturing technologies in a broad range of industries. Not all consortia were government creations but most received financial support from governments. More importantly, collaborative efforts took many different forms. Joint ventures and strategic alliances among companies in the same industry and with suppliers and subcontractors also became common near the end of the century.

While companies engaged in international alliances for generations, collaborative research efforts between overseas competitors, such as Toyota and General Motors, became increasingly common toward the end of the century. Often these international collaborations involved research to create entirely new technologies. Another trend that had earlier precedents but became increasingly common in the 1980s and 1990s was the establishment of research laboratories in foreign countries. Firms from Europe, Japan, and the U.S. established laboratories outside their countries to recruit researchers and to exploit the scientific and technological capabilities found in these markets and universities.

Increased corporate spending on university research was perhaps the defining characteristic in the organization of science and technology in the late twentieth century. Rather than hire a smattering of university professors as consultants or let small contracts to university departments, major corporations awarded multi-million dollar grants to gain access to the fruits of university research. Monsanto's $23 million grant to Harvard University in 1974 started the trend. Companies in the chemical and biotechnology industries, fields in which scientific breakthroughs are quickly translated into new products, were among the most active supporters of academic research. The trend was not limited to U.S. firms. Hoeschst of West Germany gave $50 million for research in molecular biology to Harvard's Massachusetts General Hospital. Although some university scientists rejected private funding, claiming it perverts basic research, declining state and federal support for universities had all but made industry support essential for maintaining the health and well-being of university research programs.

The shift from fundamental to product-related research and the decentralization of research within firms, the growth of national and international partnerships, the overseas expansion of R&D, and increasingly close relations between universities and corporations represented the emergence of a market-driven approach to R&D in which personal and institutional networks, rather than centralized hierarchies, defined the context of innovation to an increasing extent. The role of universities in preserving and advancing national scientific and technological capabilities appears to have increased in importance in this new organizational environment.

GLEN ASNER

Further Reading

Branscomb, L.M., Kodama, F. and Florida, R., Eds. *Industrializing Knowledge: University–Industry Linkages in Japan and the United States*. MIT Press, Cambridge, MA, 1999.

Bush, V. *Science, the Endless Frontier*. U.S. Government Printing Office, Washington D.C., 1945.

Capshaw, J.H. and Rader, K.A. Big science: price to the present. *OSIRIS*, 2nd series, 7, 3–25, 1992.

Dunning, J. Multinational enterprises and globalization of innovatory capacity. *Res. Policy*, 23, 67–88, 1994.

Dupree, A.H. *Science in the Federal Government: A History of Policies and Activities to 1940*. Harvard University Press, Cambridge, MA, 1957.

Galison, P. and Hevly, B., Eds. *Big Science: The Growth of Large-Scale Research*. Stanford University Press, Stanford, 1992.

Geiger, R.L. *Research and Relevant Knowledge: American Research Universities Since World War II*. Oxford University Press, New York, 1993.

Graham, L.R. Big science in the last years of the big Soviet Union. *OSIRIS*, 2nd series, 7, 49–71, 1992.

Hounshell, D.A. The evolution of industrial research in the United States, in *Engines of Innovation: U.S. Industrial Research at the End of an Era*, Rosenbloom, R.S. and Spencer, W.J., Eds. Harvard Business School Press, Boston, 1996.

Kevles, D.J. *The Physicists: The History of a Scientific Community in Modern America*, 3rd edn. Harvard University Press, Cambridge, MA, 1995. Originally published by Knopf, New York, 1977.

Mack, P., Ed. *From Engineering Science to Big Science: The NACA and NASA Collier Trophy Research Project Winners*. NASA History Office, Washington D.C., 1998.

Marsch, U. Strategies for success: research organization in German chemical companies and IG Farben until 1936. *Hist. Technol.*, 12, 23–77, 1994.

Meyer-Thurow, G. The Industrialization of invention: a case study from the German chemical industry. *ISIS*, 73, 268, 363–381, 1982.

Millard, A. *Edison and the Business of Innovation*. Johns Hopkins University Press, Baltimore, 1990.

Mowery, D.C. and Rosenberg, N. *Technology and the Pursuit of Economic Growth*. Cambridge University Press, New York, 1989.

Mowery, D.C. and Teece, D.J. Strategic alliances and industrial research, in *Engines of Innovation: U.S. Industrial Research at the End of an Era*, Rosenbloom, R.S. and Spencer, W.J, Eds. Harvard Business School Press, Boston, 1996.

Niosi, J. The internationalization of industrial R&D: from technology transfer to the learning organization. *Res. Policy*, 28, 107–117, 1999.

Pearce, R.D. *The Internationalization of Research and Development by Multinational Enterprises*. St. Martin's Press, New York, 1989.

Reich, L.S. *The Making of American Industrial Research: Science and Business at GE and Bell, 1876–1926*. Cambridge University Press, New York, 1985.

Rosenberg, N. and Nelson, R.R. The roles of universities in the advance of industrial technology, in *Engines of Innovation: U.S. Industrial Research at the End of an Era*, Rosenbloom, R.S. and Spencer, W.J, Eds. Harvard Business School Press, Boston, 1996.

Seidel, R.W. Accelerators and national security: the evolution of science policy for high-energy physics, 1947–1967. *Hist. Technol.*, 11, 361–391, 1994.

Servos, J.W. The industrial relations of science: chemical engineering at MIT, 1900–1939. *ISIS*, 71, 259, 531–549, 1980.

Servos, J.W. Changing partners: the Mellon Institute, private industry, and the federal patron. *Technol. Cult.*, 35, 221–257, 1994.

Siddiqi, A. *Challenge to Apollo: The Soviet Union and Space Race, 1945–1974*. NASA History Office, Washington D.C., 1998.

Smith, B.L.R. *American Science Policy Since World War II*, Brookings Institution, Washington D.C., 1990.

Swann, J.P. *Academic Scientists and the Pharmaceutical Industry: Cooperative Research in Twentieth-Century America*. Johns Hopkins University Press, Baltimore, 1988.

Varcoe, I. Cooperative research associations in British industry, 1918–1934. *Minerva*, 19, 433–463, 1981.

Westwick, P.J. *The National Labs: Science in an American System, 1947–1974*. Harvard University Press, Cambridge, MA, 2003.

Wise, G. and Whitney, W.R. *General Electric, and the Origins of U.S. Industrial Research*. Columbia University Press, New York, 1985.

P

Packet Switching

Historically the first communications networks were telegraphic—the electrical telegraph replacing the mechanical semaphore stations in the mid-nineteenth century. The network consisted of strategically located stations linked by telegraph lines. Using alphabetic codes such the five-unit Baudot code (still used in telex today), messages were relayed from station to station. This approach is known as message store-and-forward or message switching. Messages could be relayed coast-to-coast in a matter of hours or even minutes. One of the problems with message switching is that messages are highly variable in length. A short message may have to queue behind one or more long messages from other senders, and the delivery time is therefore rather variable.

Telegraph networks were largely eclipsed by the advent of the voice (telephone) network, which first appeared in the late nineteenth century, and provided the immediacy of voice conversation. The Public Switched Telephone Network allows a subscriber to dial a connection to another subscriber, with the connection being a series of telephone lines connected together through switches at the telephone exchanges along the route. This technique is known as circuit switching, as a circuit is set up between the subscribers, and is held until the call is cleared (see Telephony, Automatic Systems).

One of the disadvantages of circuit switching is the fact that the capacity of the link is often significantly underused due to silences in the conversation, but the spare capacity cannot be shared with other traffic. Another disadvantage is the time it takes to establish the connection before the conversation can begin. One could liken this to sending a railway engine from London to Edinburgh to set the points before returning to pick up the carriages. What is required is a compromise between the immediacy of conversation on an established circuit-switched connection, with the ad hoc delivery of a store-and-forward message system. This is what packet switching is designed to provide.

Packet switching is the data transmission technique whereby messages are cut up into small fixed-length pieces for routing through a data communication network, and then the pieces are reassembled into the original message at the receiving end. The term packet switching was coined in 1965 by Donald Davies, at the U.K.'s National Physical Laboratory (NPL). He proposed the approach as a solution to the burgeoning requirements for computer networks and data communication. At about the same time, another researcher, Paul Baran of the Rand Corporation in the U.S., proposed a very similar scheme, but for a highly reliable, digital, nationwide military voice network that could switch to another route when individual nodes or exchanges were destroyed by enemy action. Neither author was aware of the work of the other until a year or so later.

The technique can be applied to a wide variety of traffic types including interactive computing (as in client/server systems), file transfer, telemetry, digital telephone, or even delivery of video or audio streams. By dividing up large messages into small, fixed-length messages (packets of say, 1000 bits) and sending these like telegrams, the message handling exchanges can route each packet very simply. Being small, packets can be held in random access memory (RAM) and routed dynamically to

the next best link. Using a line of 1 megabit per second, a packet of 1000 bits takes 1 millisecond to transmit. If a packet switch can route a packet in 1 millisecond and if there are five hops between subscribers, the typical cross-network delay would be 10 milliseconds. At 50 percent loading, a 1-megabit per second line could deliver 500 packets per second and a multiprocessor switch could process thousands of packets per second. Doing this makes the packet-switched network appear as a set of direct connections between subscribers.

Message reassembly is handled at the receiving end, as is the detection of any lost packets. The technique here is for the sender to number the packets, and the receiver to request the retransmission of missing items. This happens infrequently and does not turn out to be such an onerous task, but in any case, the software for doing this is now readily available.

During 1966, Davies's team at the NPL produced a design for an actual communication network. This work was published in the proceedings of the ACM (Association for Computing Machinery) conference, held in Gatlinburg, U.S. in 1967. At that meeting a paper was presented by Larry Roberts of the Advanced Research Projects Agency (ARPA) project, describing proposals for a computer network that facilitated the sharing of resources between university and research campuses. It was a seminal meeting as the NPL proposal illustrated how the communications for such a resource-sharing computer network could be realized. During 1968 and 1969, development work proceeded in both U.S. and the U.K. The ARPA network or ARPAnet communications system was built by Bolt, Beranek & Newman Inc., and the first exchanges—interface message processors (IMPs)—were delivered in 1969. During the 1970s the network grew to some 30-plus "nodes" covering mainland U.S., with outposts in Hawaii and Europe. At the same time, a campus network was built by the NPL team to meet the laboratory's local data processing needs. Elements of the pilot NPL network first operated in 1969 and this was later expanded to link some 20 computers and 300 user terminals. The U.K. and ARPAnet networks were later connected together, and then to other experimental networks in Europe, such as the European Informatics Network, which linked France, Germany, Italy, Sweden, Switzerland and U.K.

Theoretical work was being undertaken in parallel, involving mathematical modeling and simulation into the performance of such systems. The ARPA development was underpinned by the work of Leonard Kleinrock at University of California, Los Angeles (UCLA), who had been developing such modeling techniques as a graduate student at MIT in the mid-1960s. Simulation work on packet networks was also undertaken by the NPL group.

The network research community formed the Inter-Network Working Group (INWG)—chaired by Vint Cerf—and out of this came the inter-network protocol transmission control protocol/ Internet protocol (TCP/IP, the now de facto Internet standard). The ARPA network rapidly became the focus of attention for both the computer industry and the regulated telecommunication network providers. Pressure mounted for such services in the public domain. With the momentum gathering for computer networking during the 1970s and 1980s, effort was put into public standards for such networks. In the telecommunications arena, the "regulated carriers" introduced packet services, based on the X25 standard.

Public wide area networks were characterized by the low data rates that could be supported by the telephone network infrastructure. However, the NPL work in particular had demonstrated both the benefits and the feasibility of very high data rates (around 1 megabit per second). Despite the development of digital telephone lines (2 megabits per second) the benefits of fast digital networks probably first came to the public attention in local area networks (LANs) with the design for Ethernet (10 megabits per second) by Bob Metcalf of Xerox. This design, and the competing IBM Token-ring LAN, are also packet-organized network technologies.

In the 1990s, the International Public Record Carriers started to deploy even faster digital communication systems based on packet technology (frame relay and asynchronous transfer mode (ATM), the latter at 155 megabits per second). The term "Internet" was coined and Tim Berners-Lee conceived the notion of the World Wide Web. This popular technology outstripped the ability of the regulated carriers to satisfy demand, and the essentially unregulated Internet spread like wildfire, fueled by the burgeoning home PC market. At the turn of the century, packet switching was embedded in so much that it was accepted as daily experience. Home users could connect to the Internet at near-megabit rates using ADSL (asynchronous digital subscriber line), and gigabit Ethernet was introduced. In optical communication fiber networks, optoelectronic switches route optical signals. Current optical–electronic–optical

(O–E–O) conversion will not be able to support the terabit-per-second capacities that will be needed in the near future. Future high-speed packet switching will only be realized by all-optical (photonic) networks.

See also **Computer Networks; Electronic Communications; Internet; Telephony, Automatic Systems; Telephony, Digital; World Wide Web**

ROGER SCANTLEBURY

Further Reading

Abbate, J. *Inventing the Internet*. MIT Press, Cambridge, Massachusetts, 2000.

Campbell-Kelly, M. Data communications at the National Physical Laboratory (1965–1975). *Ann. Hist. Comput.*, 9, 3/4, 221–247, 1988.

Gillies, J. and Cailliau, R. *How the Web was Born*. Oxford University Press, Oxford, 2000.

Norberg, A.L. and O'Neill, J.E. Improving connections among researchers: the development of packet-switching computer networks, in *Transforming Computer Technology*. Johns Hopkins University Press, Baltimore, 1986.

Yates, D.M. *Turing's Legacy: A History of Computing at the NPL, 1945–1995*. Science Museum, London.

Particle Accelerators: Cyclotrons, Synchrotrons, and Colliders

Particle accelerators, or "atom smashers" as they are popularly known, are devices that produce concentrated beams of charged particles of very high energy. These beams have been used to study nuclear and atomic structure (x-ray crystallography), to create radioactive isotopes, and to irradiate cancerous tumors with x-rays. After World War II they constituted the essential infrastructure for the new field of high-energy physics. As accelerator power climbed to ever-higher energies, the laboratories where these microscopes probed to the heart of matter were transformed into huge industrial-scale centers of "big science." Particle accelerators became particle factories supporting multidisciplinary teams of researchers whose size has increased from less than a dozen in the 1960s to as many as 1500 by the end of the twentieth century.

In the early 1930s a number of attempts were made to produce an energetic beam of charged particles under controlled conditions. One of the simplest devices was that devised by John Cockroft and Ernest Walton working in Ernest Rutherford's famous Cavendish Laboratory in Cambridge. By charging a bank of capacitors in parallel at low potential and then discharging them through a load resistor to develop a high potential, they could multiply voltage by a factor of 4. The British scientists attained a steady output of about 500 kilovolts. This was used to accelerate protons (positively charged hydrogen ions) that were then smashed into atoms in a metallic target. Another scheme was the electrostatic generator developed by Robert J. Van de Graaff working at Princeton University. Ions produced by a corona discharge from needle points were transported by motor-driven insulating belts to two spherical conductors mounted on insulating rods. There the belts were discharged into the terminals. A potential difference of up to 1.5 million volts could be accumulated on the spheres, limited only by voltage breakdown, and sparking between them. This approach was so successful that Van de Graff and others formed their own company after the war to commercialize their device.

In these generators, charged particles picked up energy by falling through a large voltage difference. An alternative idea, also considered at the time, was to have the charged particles gain energy in several small steps. In this way, "the high-voltage energy would be accumulated on the particles, not on the apparatus" (Heilbron and Seidel, 1989). It was Ernest O. Lawrence, working at the Radiation Laboratory at the University of California in Berkeley, along with his graduate student M. Stanley Livingston, who first successfully applied this concept in the autumn of 1931, accelerating protons to over 1 million volts in a cyclotron. Lawrence won the 1939 Nobel Prize in Physics for his invention.

The cyclotron comprised two hollow half-cylinders or D's, with a small gap between them. A magnetic field was applied perpendicularly to the D's. Ions were injected at the center, and were given a small kick by a radio oscillator as they crossed the gap. They described a circular trajectory in the magnetic field, incrementally increasing their energy (say, by 10 kilovolts), and the radius of their path, each time they crossed the gap (say, 100 times in all). Tracing a spiral as they moved from the center of the D's to the circumference they thus emerged with 1 million volts of energy (in this case).

The key to increasing energy lay in the size of the D's. The first experimental setup on which Livingston demonstrated the feasibility of the idea could be fitted in the palm of one's hand. Within the decade it was followed by the 27-inch (69-centimeter) D, the 60-inch (152-centimeter) D, and then the huge 184-inch (467-centimeter) D; this measure being the diameter of the magnet face.

This machine, largely funded by the Rockefeller Foundation at a cost of $1.4 million in 1940, was designed to reach energies higher than 100 million volts, and required its own building to house it.

Cyclotron energies were restricted by the fact that, as the velocity of the particles approached that of the speed of light, the mass changed in accordance with Einstein's relativistic principles, and the orbital frequency of the particles changed with it. The principles on which the cyclotron was based thus broke down. This limitation on achievable energy was removed with the discovery of phase stability in 1945. The implementation of this technique depended on the whether it was applied to a proton or an electron accelerator. In the case of protons, the decreasing orbital frequency was compensated for by increasing both the strength of the magnetic field and the frequency of the accelerating voltage. Phase stability allowed engineers to do away with the huge pole faces required in a high-energy cyclotron. The particle beams circulated in evacuated beam pipes surrounded by magnets, radio-frequency generators and power supplies. Economic considerations were the only remaining constraint on particle energy, and the field of high-energy physics was born.

Many innovations have been exploited to increase accelerator energy while containing costs. Strong focusing, discovered in the U.S. in 1952, was an ingenious technique for limiting the cross-section of the particle beam, and therefore of the beam tube in which it circulated. The size (and therefore the cost) of the magnets confining the beam was thus sharply reduced. Colliding beams of particles traveling in opposite directions is another important way of increasing the energy available for doing physics. The first practical demonstration of this principle occurred at the European Laboratory for Particle Physics (CERN, near Geneva) in 1971. The beams were produced in two different intersecting storage rings (the ISR machine). Drawing on this experience, CERN scientists won the Nobel Prize for Physics in 1984 for colliding beams of opposite charge circulating in the same ring. The use of superconducting magnets became practicable in the 1980s, and these have opened yet another cost-reducing path to higher energy.

Most laboratories use circular accelerators. Linear accelerators have also been built, notably the 3.2-kilometer-long machine authorized in 1959 near Stanford University in California. Practicable length is, however, a limitation on energy, and linacs are generally used as injectors into ring systems. Today these are gigantic and very expensive. The new machine under construction at CERN, the large hadron collider, will be housed in a tunnel 27 kilometers in circumference. In 2001 its cost to completion was estimated to be about $1,800 million. The superconducting super collider (SSC), cancelled by the Clinton administration in 1993 when its costs began to creep above $10,000 million, was designed to reach 20 million million volts in a 85-kilometer-long subterranean tunnel in Texas.

Only some governments have been able to afford tools of this kind for such esoteric research, and they have done so for reasons of state. CERN, established in 1954 by 12 European governments who pooled their resources, combined a determination to keep nuclear scientists in Europe with foreign policies sympathetic to the construction of a European community. In the U.S., particle accelerators were identified with world scientific and technological leadership, national prestige, and the "peaceful atom," and used to promote international scientific exchange behind the Iron Curtain. Correlatively, some would argue, the demise of the SSC is simply one more response to the collapse of the Soviet bloc.

See also **Cancer: Radiation Therapy; Particle Accelerators, Linear**

JOHN KRIGE

Further Reading

Galison, P. and Hevly, B. Eds. *Big Science. The Growth of Large-Scale Research*. Stanford University Press, Stanford, 1992.

Heilbron, J.L., Seidel, R.W. and Wheaton, B.R. Lawrence and his laboratory: nuclear science at Berkeley. *LBL News Magazine*, 6, 3, 1981.

Heilbron, J.L. and Seidel, R.L. *Lawrence and His Laboratory. A History of the Lawrence Berkeley Laboratory*, vol. I. University of California Press, Berkeley, 1989.

Hermann, A., Krige, J., Mersits, U. and Pestre, D. *History of CERN. Vol. I: Launching the European Organization for Nuclear Research*. North Holland, Amsterdam, 1987.

Livingston, M. S. *Particle Accelerators. A Brief History*. Harvard University Press, Cambridge, MA, 1969.

Riordan, M. The demise of the superconducting super collider. *Phys. Perspect.*, 2, 411–425, 2000.

Westfall, C. and Hoddeson, L. Thinking small in big science: the founding of Fermilab. *Technol. Cult.*, 37, 457–492, 1996.

Useful Websites

CERN, European Organization of Nuclear Research: http://www.cern.ch

Particle Accelerators, Linear

A linear accelerator or linac, as it is commonly called, is a device that uses an oscillating electric field to accelerate charged particles (atomic or subatomic particles) in a straight line. The final energy of the charged particles is achieved by repeated acceleration on a linear path through energy steps.

The high-energy particle beams produced are used in elementary particle physics research, as neutron sources for research and materials inspection, and in radiation therapy for treatment of deep cancers (either as direct electron treatment, or from x-rays produced when high-energy electrons strike a target). Accelerated electrons or x-rays, both of which also cause ionization, can also be used to irradiate food against insect pests and pathogenic bacteria.

The 1920s saw not only the first formal proposal of a linear accelerator by the Swedish scientist G. Ising, but also the first report of a working one by the Norwegian R. Wideröe. It was in 1928 when Wideröe described his experiments with a machine consisting of three consecutive cylindrical electrodes, whereby an alternating electric field was applied between the central electrode and the two side ones. The frequency of alternation was such that, during the time in which the potassium (K^+) and sodium (Na^+) ions he used traversed the central electrode, the electrode potentials were reversed. Actual acceleration took place when particles traveled from one electrode to the next. This is how, with the central electrode shielding the particles for the length of time that the field would be decelerating, ions reached a final energy twice the energy obtained from a single transversal of the field. The radio-frequency voltage of 25 kilovolts used sped ions up to energies of 50 kiloelectron volts of energy. During the following two decades, several proposals and experimental devices came to light in the attempt to accelerate particles to higher energies. From what we have said, it is clear that high-voltage machines can be considered to be early linear accelerators. In these devices, a high-voltage generator provides a potential difference into an accelerating chamber, and the particle beam acquires energies of the order of the voltage delivered by the generator. In 1932 the British John D. Cockcroft and the Irish Ernest T. S. Walton used a linear machine of this kind to accelerate protons to 500 kiloelectron volts and disintegrate lithium nuclei at the Cavendish Laboratory at Cambridge University.

During the 1930s, a group of American physicists worked in Berkeley, California, developing and improving Wideröe's idea. The work of Ernest O. Lawrence and D. H. Sloan culminated in a machine that accelerated mercury (Hg^+) ions to energies of up to 1.26 megaelectron volts. The design was based on 30 consecutive cylindrical metallic electrodes of increasing length—note that particle velocity increases as it travels through the accelerator—placed along the axis of a vacuum chamber. These electrodes were connected alternately to the positive and negative polarities of a 42-kilovolt alternating current voltage generator. Obviously, the longer the accelerator, the higher the energies obtained. The later design of 36 drift tubes, by Sloan and W. M. Coates, achieved final energies of 2.85 megaelectron volts. Nevertheless, heavy ion linear accelerators seemed less useful in the field of nuclear physics than circular accelerators. The investment in the field of radar and radio communication during World War II allowed the development of power generators in the megawatt power range and in the gigahertz frequency range. This was what physicists needed for successfully accelerating lighter particles and building practical linear accelerators with the energies required for nuclear physics studies. At Stanford University, American William W. Hansen and his research group proceeded toward the construction of an electron linac, a project conceived before the war, based on Russell and Sigurd Varian's klystron (which generated the microwave power necessary to accelerate electrons, and was used in World War II in radar aboard aircraft), and completed in 1947. The machine, later known as Stanford Mark I, was 2.7 meters long, powered by a single magnetron that yielded 0.9 megawatts, and provided electrons with an energy of around 4.5 megaelectron volts. A short time later, it was extended to 4.3 meters long and energies of 6 megaelectron volts were obtained. At the same time, at the University of California, design studies under the direction of American Luis W. Alvarez led to the construction of a proton linear accelerator. The surplus radar components available from the war were adopted to power a machine 12.2 meters long. The Berkeley 32-mega-electron-volt proton linear accelerator was in full operation by 1947, two years after its construction began.

D.W. Fry's group at the Telecommunications Research Establishment (TRE), Great Malvern, U.K, worked more or less simultaneously with Hansen on a linear electron accelerator, but with limited knowledge of Hansen's work until mid-

1947. Fry's accelerator was driven by a magnetron, and produced a beam of 0.5 megaelectron volt electrons towards the end of 1946. The British linear accelerator was soon adopted by hospitals for cancer therapy, with the first patient treated in August 1953 at Hammersmith Hospital in London. In the U.S., Henry Kaplan, working with Edward Ginzton of the Stanford Physics Department, worked on medical applications. The Stanford medical linear accelerator was completed in 1955.

Several developments and improvements made work at higher energies in modern linear accelerators possible. In each case, different problems appeared to be related to the shape of electromagnetic fields, particle dynamics, power requirements, cavities, and defocusing. Well-known examples are the 91-meter Stanford Mark III, operating in 1964 at 1.2 gigaelectron volts energy; and the largest linac built, the 3.2-kilometer machine at the Stanford Linear Accelerator Center (SLAC), with emerging electron energies of more than 20 gigavolt electrons in 1967. During the 1960s and the 1970s, ideas such as the linear induction accelerator, where inductive electric fields are used, contributed to progress in the pursuit of higher energies and new challenges. The invention of the low-energy radio-frequency quadrupole (RFQ) linear accelerator, where the electric fields are produced by the radio-frequency fields applied to four electrodes collinear with the beam axis, was proposed by the Russians I. M. Kapchiski and V. A. Teplyakov and became one of the most important improvements.

The construction and development of linear accelerators has continued today, not only because of interest in electron and ion optics, space science, industrial applications, and thermonuclear fusion, but also for medical purposes. An example of this is the Los Alamos Meson Physics Facility (LAMPF), an 800-megaelectron-volt proton linear accelerator sponsored by the National Cancer Institute. There is no doubt that the excellent collimation of the emergent beam, as opposed to the spreading that results from circular accelerators, is one of the most attractive advantages that linear accelerators offer for such purposes. Nevertheless, the most common use of proton linear accelerators is as injectors for high-energy machines. Electron linear accelerators have an advantage over circular accelerators however in that they present no practical limitation. A good example of future machines of this kind is the worldwide collaboration compact linear collider (CLIC), a linear accelerator planned by CERN. Nearly 40 kilometers long, future lepton machines will probably be linear colliders like the one proposed by CERN. Powered from a drive beam with a high frequency of 30 gigahertz, the CLIC is thought to cover a center-of-mass energy range for electron–positron collisions of 0.5 to 5 teraelectron volts.

See also **Cancer, Radiation Therapy; Particle Accelerators: Cyclotrons, Synchrotrons, Colliders**

PEDRO RUIZ CASTELL

Further Reading

Ginzton, E.L. and Nunan, C.S. History of microwave electron linear accelerators for radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 11, 205–216.

Lapostolle, P.M. and Septier, A.L., Eds. *Linear Accelerators*. North Holland, Amsterdam, 1970.

Livingston, M.S. and Blewett, J.P. *Particle Accelerators*. McGraw-Hill, New York, 1962.

Livingston, M.S. *Particle Accelerators: A Brief History*. Harvard University Press, Cambridge, MA, 1969.

Pierce, J.R. *Theory and Design of Electron Beams*. Van Nostrand, New York, 1949.

Reiser, M. *Theory and Design of Charged Particle Beams*. Wiley, New York, 1994.

Scharf, W.H. *Particle Accelerators: Applications in Technology and Research*. Research Studies Press, Taunton, 1989.

Wiedeman, H. *Particle Accelerator Physics: Basic Principles and Linear Beam Dynamics*. Springer Verlag, Berlin, 1993.

Useful Websites

AccSys Technology, Inc.: http://www.accsys.com

CERN, European Organization of Nuclear Research: http://www.cern.ch

Stanford Linear Accelerator Center: http://www.slac.stanford.edu

Personal Computer, *see* **Computers, Personal**

Personal Stereo

The scaled-down cassette tape player represents not only one of the most successful audio products of the twentieth century but also a drastic change in the way we listen to music. The personal stereo has become a universal product that can be found in every corner of the globe. It has brought high-quality reproduction of sound into every field of human activity.

The concept of the personal stereo can be traced to two great electrical manufacturers: Sony of Japan and Philips of the Netherlands. Both of these companies created a great new market for electrical goods by reducing the size of their products. Key employees in both companies shared

a passion for music and a goal of perfect reproduction of sound.

Akio Morita and Masuru Ibuka formed a company in 1946 to make a variety of electrical testing devices and instruments, but their real interests lay in music and they decided to concentrate on audio products. They renamed their company Sony (derived from "sonic") to emphasize this strategy. A pioneer in developing the pocket-sized transistor radio, Sony penetrated the American market with small, fully transistorized television receivers.

The original marketing strategy for manufacturers of all mechanical entertainers had been to put one into every home. This was the goal for Edison's phonograph, the player piano, the Victor Talking Machine Company's Victrola and the radio receiver. However, Sony and other Japanese manufacturers found out that if a product was small enough and cheap enough, two or three might be purchased for home use. This was the marketing lesson of the transistor radio that was successfully applied to televisions and tape players.

The personal stereo was the result of the convergence of two technologies: the transistor, which enabled miniaturization of electronic components, and the compact cassette—a worldwide standard for magnetic recording tape. The latter was devised by Philips as a replacement for the cumbersome reels used in tape recorders. Users found threading tape around the reels troublesome. The size of the reels (and the power requirements to turn them) made it difficult to reduce the size of tape recorders. Philips' cassette was one of many similar innovations in the 1960s based on the tape cartridge concept already in use in film cameras. At this point there was no idea that a smaller tape recorder would have applications in musical entertainment; the goal was to develop small, portable recorders to be used as dictating machines. Philips' executive Lou Ottens played an important part in this project, applying the descriptive adjective "compact" to the cassette.

Masuru Ibuka of Sony initiated the research project that led to the Walkman personal stereo. He wanted to be able to listen to high-fidelity recorded sound wherever he went and instructed his team to produce a player small enough to fit inside his pocket. (Another group of Sony engineers was working on a video recording cassette that could also fit into Ibuka's pocket.) The Walkman was based on a systems approach that made use of advances in several unrelated areas, including improvements in magnetic tape, integrated circuits, and new types of batteries (notably the nickel–cadmium combination, which offered higher output in smaller sizes). The problem of reducing the size of the loudspeaker without serious deterioration of sound quality blocked the path to very small cassette players. Sony's engineers produced a very small dynamic loudspeaker using plastic diaphragms and lighter materials for the magnets. These were incorporated into tiny stereo headphones.

The Sony Soundabout portable cassette player was introduced in 1979. It was initially treated as a novelty in the audio equipment industry. Sony's engineers, working under the direction of Kozo Ohsone, reduced the size and cost of the machine. In 1981 the Walkman II was introduced. It was 25 percent smaller than the original version and had 50 percent less moving parts. It took about two years for Sony's Japanese competitors, including Matsushita, Toshiba and Aiwa, to bring out portable personal stereos. Such was the popularity of the device that any miniature cassette player was called a Walkman, irrespective of the manufacturer. In 1986 the term entered the *Oxford English Dictionary*.

Constant innovation added new features to the personal stereo: Dolby noise reduction circuits were added in 1982 and rechargeable batteries were introduced in 1985. The machine grew smaller and smaller until it was hardly bigger than the compact cassette it played.

Within two years of the introduction of the compact disk in 1982, Sony had brought out a portable player named the Discman.

Ohsone led the development team that had to overcome the considerable challenges of reducing vibration in the unit (which disturbed the optical reading of microscopic lines of data) and reducing the size of the laser reader. Like its cassette counterpart, the Walkman technology was systematically improved while its size and price was continually reduced.

The size of the market for personal stereo systems ensured that many new recording technologies developed in the 1990s would be reduced in size and offered with earphones. This was the case for digital audio tape (DAT), Philips' digital compact cassette (DCC), and Sony's minidisc (MD). All these technologies came with the vital advantage of a recording capability—the major commercial consideration in competing with magnetic tape units.

The minidisc has been the most successful digital version of the personal stereo recorder. It employs the same optical technology as the compact disk

but in a smaller size. It contains enough buffer memory to overcome the skipping of tracks that caused problems with the Discman.

The ubiquitous Walkman has had a noticeable effect on the way that people listen to music. The sound from the headphones of a portable player is more intimate and immediate than the sound coming from the loudspeaker of the home stereo; the listener can hear a wider range of frequencies and more of the lower amplitudes of music, while the reverberation caused by sound bouncing off walls is reduced. The listening public have become accustomed to the Walkman sound and expect it to be duplicated on commercial recordings. Recording studios that once mixed the balance of their master recordings to suit the reproduction characteristics of car or transistor radios now mix them for Walkman headphones.

See also **Audio Systems**

ANDRE MILLARD

Further Reading

Abbate, J. The electrical century: getting small—a short history of the personal comp. *Proc. IEEE*, 87, 1695–1698, 1999.

Klein, L. Happy Anniversary, Sony Walkman! *Radio Electron.*, 60, 72–73, 1989.

Lyons, N. *The Sony Vision*. Crown, New York, 1976.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, New York, 1995.

Morita, A., Reingold, E. and Shimomura, M. *Made in Japan: Akio Morita and Sony*. Hutton, New York, 1986.

Read, O. and Welch, W.L. *From Tin Foil to Stereo: Evolution of the Phonograph*. Sams, Indianapolis, 1976.

Schiffer, M. *The Portable Radio in American Life*. University of Arizona Press, Tucson, 1991.

Pest Control, Biological

Insect outbreaks have plagued crop production throughout human history, but the growth of commercial agriculture since the middle of the nineteenth century has increased their acuteness and brought forth the need to devise efficient methods of insect control. Methods such as the spraying of insecticides, the application of cultural methods, the breeding of insect-resistant plants, and the use of biological control have increasingly been used in the twentieth century. Traditionally limited to checking the populations of insect pests through the release of predatory or parasitic insects, biological control now refers to the regulation of agricultural or forest pests (especially insects, weeds and mammals) using living organisms. It also includes other methods such as the spraying of microbial insecticides, the release of pathogenic microorganisms (fungi, bacteria or viruses), the release of male insects sterilized by radiation, the combination of control methods in integrated pest management programs, and the insertion of toxic genes into plants through genetic engineering techniques. Biological control is also directed against invasive foreign species that threaten ecological biodiversity and landscape esthetics in nonagricultural environments.

The Chinese are known to have used natural enemies to control insect pests as far back as the ninth century and European naturalists and agriculturists were proposing the use of entomophagous insects in the eighteenth century. Biological control only seriously took off at the end of nineteenth century, however, following a successful insect control campaign by the U.S. government. In 1888, Charles Valentine Riley, chief entomologist of the U.S. Department of Agriculture, sent one of his field entomologists, Albert Koebele, to Australia to search for natural enemies of the cottony cushion scale, an insect pest accidentally introduced in the citrus groves of California in 1869. Koebele found a lady beetle of the genus *Vedalia* that fed on the scale. Specimens of the beetle were sent to California, then bred and distributed throughout the state. Suppression of the scale followed within two years of the beetle's introduction.

This achievement incited other countries to initiate their own campaigns of biological control. Australia, Canada and the U.S. became especially proficient, as these countries were the most affected by the accidental introduction of foreign insects; insect pests rapidly reached outbreak levels in environments that contained none of the adverse conditions that would have normally limited their multiplication. Under the leadership of Riley's successor, Leland O. Howard, the U.S. Bureau of Entomology (USBE) set up laboratories for breeding parasites of destructive insects such as the gypsy moth, the Japanese beetle and the European corn borer, and it posted entomologists abroad to collect natural enemies. At the University of California, Harry S. Smith, the entomologist who coined the expression biological control in 1919, organized a Division of Beneficial Insect Investigations where he trained generations of practitioners. In England, the Imperial Bureau of Entomology established the Farnham House Laboratory (renamed the Commonwealth Institute of Biological Control (CIBC) in 1948) to coordinate and conduct biological control campaigns in the British colonies and dominions.

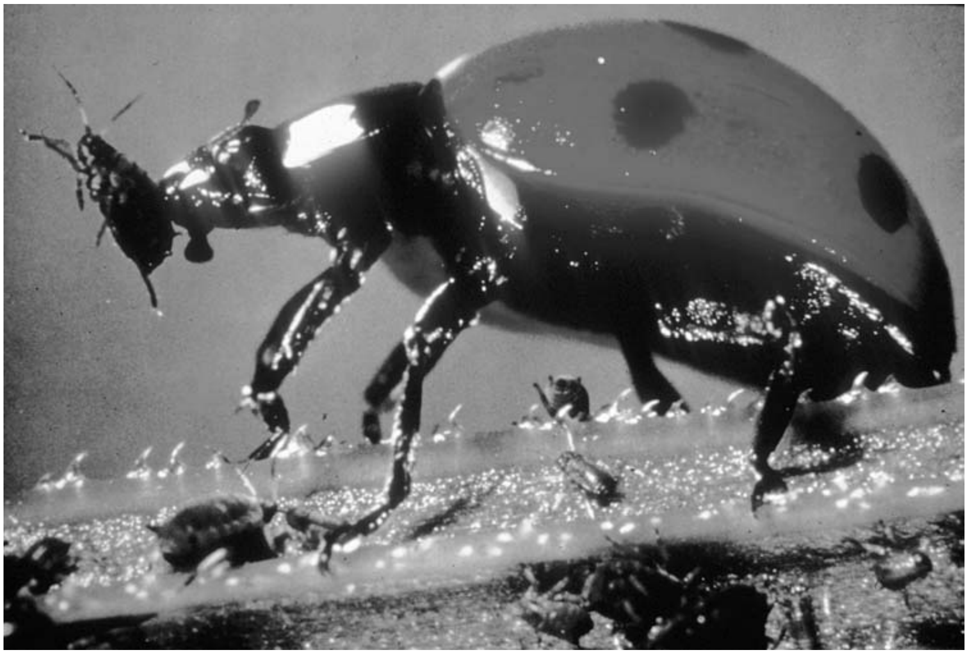


Figure 1. Sevenspotted Lady Beetle, Coleoptera:Coccinellidae, a predator of aphids. [*Photography : courtesy of Daniel Coderre, Laboratoire de lutte biologique, Université du Québec à Montréal.*]

However, the instantaneous and spectacular success of *Vedalia* was rarely repeated, and support for biological control dwindled in the interwar period. The introduction of synthetic organic insecticides during World War II significantly transformed applied entomology and further challenged the relevance of biological control. Their extreme toxicity proved to be an important drawback, however, and it soon became evident that they were not the panacea that many thought they would be.

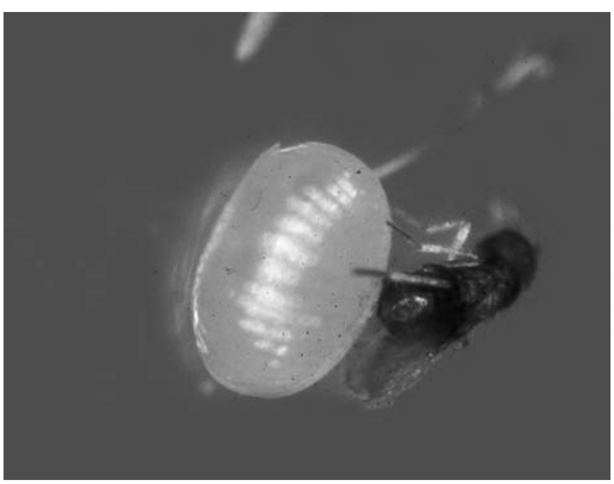


Figure 2. Trichogramma, Hymenoptera: Trichogrammatidae, a parasitoid wasp. Larvae develop inside the host egg, eventually destroying it. [*Photography : courtesy of Daniel Coderre, Laboratoire de lutte biologique, Université du Québec à Montréal.*]

Microbial control—the use of pathogenic organisms—offered an attractive alternative. During the 1930s, the fortuitous spread of two insect diseases demonstrated the rapidity and efficiency of microorganisms in controlling outbreaks: a bacterium drastically reduced the Japanese beetle population in the U.S. and a polyhedral virus ended the European spruce sawfly outbreak in Canada. Advances in the microbiological sciences and research on diseases of beneficial insects like silkworms and bees provided the knowledge to propagate pure cultures of entomopathogens and to maintain their virulence in the field. In the 1950s, the commercialization of a microbial insecticide using a toxin derived from the *Bacillus thurigiensis* (Berliner) bacteria paved the way for the aerial spraying of large forest areas in Canada and of agricultural fields in the U.S. At the time of writing, *Bacillus thurigiensis* toxin genes are inserted directly into crop plants. Although selective to certain insect pests, microbial control is not self-sustaining, and certain entomologists do not consider it a method of biological control.

Integrated control is another technique that entomologists popularized after World War II. As in biological control, integrated control relies on natural mortality factors, but it also involves the spraying of selective chemical insecticides that, applied in a timely manner, avoid inhibiting the activities of natural enemies. The timing depends on an economic threshold: microeconomic and population dynamic models are used to determine when crop damage is expected to surpass a tolerable level of loss. During the 1940s, Alison D. Pickett in eastern Canada and Harry H. Smith in California had already devised spraying programs that encouraged the survival and multiplication of natural enemies. After chemical insecticides came under public scrutiny during the 1960s, integrated control and its successor, integrated pest management, became widely accepted in policy and agricultural circles.

A characteristic of the growth of biological control in the twentieth century was the international cooperation of national and individual actors. From the outset, the USBE and the CIBC encouraged entomologists from different countries to collect and exchange specimens of natural enemies and information on the ecology of insects. With comparatively few cases of agricultural crops damaged by the introduction of foreign insects, biological control in Europe revolved mainly around the utilization of native arthropods and vertebrates. However, the need to introduce natural enemies from foreign environments eventually led European entomologists to create the International Commission for Biological Control in 1955. Founded under the aegis of the International Union of Biological Sciences, the Commission (renamed the Organization in 1965) served Western Europe, the Mediterranean basin and the former colonies of France and Belgium. In 1971, it enlarged its geographical scope and became a worldwide organization for the identification, collection and distribution of insects.

See also **Crop Protection, Spraying; Pesticides**

STÉPHANE CASTONGUAY

Further Reading

Evenden, M.D. The laborers of nature. Economic ornithology and the role of birds as agents of biological pest control in North American agriculture, ca. 1880–1930. *Forest Conserv. Hist.*, 39, 172–183, 1995.

Garcia, R., Caltagirone, L.E. and Gutierrez, A.P. Comments on a redefinition of biological control. *Bioscience*, 38, 10, 692–694, 1988.

Hagen, K. A history of biological control, in *History of Entomology*, Smith, R.F., Mittler T.E., and Smith, C.N., Eds. Annual Reviews, Palo Alto, 1973.

Huffaker, C.B., Eds. *New Technology of Pest Control*. Wiley, New York, 1980.

Kingsland, S.E. *Modeling Nature: Episodes in the History of Population Ecology*. University of Chicago Press, Chicago, 1985.

Marchal, P. *Les Sciences Biologiques Appliquées à l'Agriculture et la Lutte Contre les Ennemis des Plantes aux États-Unis [Biological Sciences Applied to Agriculture]*. Librairie l'Homme, Paris, 1916.

Palladino, P.S. *Entomology, Ecology and Agriculture: The Making of Scientific Careers in North America, 1885–1985*. Harwood, Amsterdam, 1996.

Perkins, J.H. *Insects, Experts, and the Insecticide Crisis: The Quest for New Pest Management Strategies*. Plenum Press, New York, 1982.

Sawyer, R. *To Make a Spotless Orange: Biological Control in California*. Iowa State University Press, Ames, 1996.

Smith, H.S. On some phases of insect control by the biological method. *J. Econ. Entomol.*, 12, 288–292, 1919.

Thompson, W.R. *The Biological Control of Insect and Plant Pests: A Report on the Organisation and Progress of the Work of Farnham House Laboratory*. King's Printer, London, 1930.

Pesticides

A pesticide is any chemical designed to kill pests and includes the categories of herbicide, insecticide, fungicide, avicide, and rodenticide. Individuals, governments, and private organizations used pesticides in the twentieth century, but chemical control has been especially widespread in agriculture as farmers around the world attempted to reduce crop and livestock losses due to pest infestations, thereby maximizing returns on their investment in seed, fuel, labor, machinery expenses, animals, and land.

Until the twentieth century, cultural pest control practices were more popular than chemicals. Cultural methods meant that farmers killed pests by destroying infested plant material in the fields, trapping, practicing crop rotation, cultivating, drying harvested crops, planting different crop varieties, and numerous other techniques. In the twentieth century, new chemical formulations and application equipment were the products of the growth in large-scale agriculture that simultaneously enabled that growth.

Large scale and specialized farming provided ideal feeding grounds for harmful insects. Notable early efforts in insect control began in the orchards and vineyards of California. Without annual crop rotations, growers needed additional insect control techniques to prevent build-ups of pest populations. As the scale of fruit and nut production increased in the early decades of the century, so too did the insect problem.

In the early 1900s, chemical control became one of two lines of research for entomologists. The first, biological control, was the search for natural enemies of the insect pests. Although American entomologists scored a noted success in controlling the cottony cushiony scale (an insect that devastated California's citrus crops) with Australian ladybugs in the late 1800s, most biological control efforts were not so successful. Chemical control was also a mix of failure and success. Farmers applied solutions containing arsenic, sulfur, and mercury to rid their crops and livestock of boll weevils, coddling moths, and ticks that carried Texas fever. Some farmers applied an inappropriate chemical or the wrong amount. Many, however, urged on by manufacturers, scientists, and their own need to maximize yields, still had confidence in chemical solutions.

The 1930s and 1940s were crucial decades for the development of pesticides. In 1939, Paul Mueller of the Geigy Company (Switzerland) developed an effective insect killer, dichloro-diphenyl trichloroethane (DDT). American armed forces used it to reduce the number of noncombat related casualties to record lows for soldiers and civilians, most notably controlling an outbreak of typhus in Naples in 1943 by killing body lice, the disease vector. U.S. War Research Service scientists experimented with 2, 4 dichlrophenoxyacetic acid (2,4-D), a synthetic hormone used in the 1930s to promote uniform ripening of fruit. This growth regulator, when administered in large doses, actually killed plants by stimulating them to grow themselves to death. As World War II ended, the widespread use of DDT and experiments with 2,4-D were signal accomplishments of American industry and government.

After the war, the U.S. Department of Agriculture (USDA), the agricultural extension service, and manufacturers promoted the use of pesticides among farmers and other groups including local governments, businessmen, and household decision makers. Armed with powerful new killers, Americans attempted to reduce or even eradicate pest species on golf courses, lawns, farms, and in homes, lawns, offices, restaurants, warehouses, and grain elevators. Agricultural Research Service leaders promoted fire ant eradication in the southern states, while cooperative extension experts in the Midwest promoted the elimination of various fly species. Similarly, farmers and governments around the world turned to pesticides, notably on plantations dedicated to single crops and to kill mosquitoes that carried malaria. Pesticides became popular because people liked quick, easy, and inexpensive pest control.

During the 1960s and 1970s the widespread optimism about the value of pesticides waned, especially in developed nations. In 1946, scientists cautioned farmers not to use DDT and other chlorinated hydrocarbon insecticides directly on dairy animals or beef cattle immediately before they were sold, since these chemicals were stored in fat and ended up in food products. By the late 1950s the U.S. Food and Drug Administration inspected food products and compelled farmers who sold products with residual DDT to dump their milk. In the 1960s, municipal governments aggressively used DDT to destroy the insect that carried Dutch elm disease across the U.S., but the DDT also killed songbirds, provoking public concern about the wisdom of using pesticides. Farmers and ranchers found that some plant and insect species became resistant or were already resistant to certain chemicals. Once farmers suppressed a targeted species, resistant species filled the ecological vacuum. In 1962 Rachel Carson's critique of pesticides, *Silent Spring*, attracted the attention of entomologists and the public. Carson's clear explanations and moving prose showed the harmful unanticipated consequences of chemical use by focusing on ecosystems. Concerns about the health of ecosystems and wildlife, not human health, prompted the public discussion that led to the U.S. ban of DDT in 1972. In the 1970s and 1980s, American Vietnam veterans and scientists claimed that exposure to trichlorophenoxyacetic acid (2,4,5-T), a herbicide and defoliant known as Agent Orange that included dioxin, caused illness.

Critics of the high economic and potential health costs of pesticides practiced organic farming. Although the definition of organic was a moving target, depending on USDA policy and pressure from retailers and producers, the idea of pesticide-free products was attractive to a minority of consumers and producers. Organic producers substituted labor and fuel costs for pesticides, hoping to capitalize on high-return niche markets of affluent or environmentally concerned citizens.

By the end of the twentieth century, pesticides were essential in agriculture and industry, despite some skepticism about their value. In many places in the world, chlorinated hydrocarbon insecticides continued to play a valuable role in controlling disease vectors, saving millions of lives. Herbicides played a role in reducing child labor in developing nations, allowing more children to attend school. Many industrial and agricultural chemical users of the late 1900s practiced a balance of chemical,

cultural, and environmental controls called integrated pest management (IPM) to reduce reliance on chemicals. Genetically modified organisms (GMOs), such as roundup-ready soybeans, became a significant portion of crops grown, although some nations refused to import GMO crops because of fears over unknown consequences and protectionist trade policies. Fields planted with roundup-ready plants could be sprayed with herbicide, killing weeds but preserving crops. Corn with the *Bacillus thuringensis* gene spliced into its DNA infects the European corn borers that eat the corn. These practices actually reduced the need for chemical insecticide. Consumers, producers, distributors, and retailers all found uses for pesticide in the twentieth century, although they climbed steep learning curves in adopting and adapting it.

See also **Agriculture and Food; Chemicals; Farming, Agricultural Methods; Pest Control, Biological**

J. L. ANDERSON

Further Reading

Dunlap, T.R. *DDT: Scientists, Citizens, and Public Policy*. Princeton University Press, Princeton, NJ, 1983.

Russell, E. *War and Nature: Fighting Humans and Insects with Chemicals from World War I to Silent Spring*. Cambridge University Press, Cambridge, 2001.

Schlebecker, J.T. *Whereby We Thrive: A History of American Farming, 1607–1972*. Iowa State University Press, Ames, 1975.

Sharrer, G.T. *A Kind of Fate: Agricultural Change in Virginia, 1861–1920*. Iowa State University Press, Ames, 2000.

Stoll, S. *The Fruits of Natural Advantage: Making the Industrial Countryside in California*. University of California Press, Berkeley, 1998.

Photocopiers

The photocopier, copier, or copying machine, as it is variously known, is a staple of modern life. Copies by the billions are produced not only in the office but also on machines available to the public in libraries, copy shops, stationery stores, supermarkets, and a wide variety of other commercial facilities.

Over the years, various processes have been employed. By far the most common type of photocopier today is the electrostatic, or xerographic. It is the type most people are familiar with, and arguably that with which most people associate the term photocopier. It was the electrostatic process that revolutionized copying as a part of everyday life. The modern photocopying era began in 1960 with the introduction of the Xerox 914, the first push-button, plain-paper copier. Within one year, sales doubled, and *Fortune* magazine called the 914 "the most successful product ever marketed in America."

While "photocopying" to most people means copying with the degree of the quality and speed that has existed since the 1960s, copying technology in a rudimentary form can be dated to the pantograph of the seventeenth century. The standard of the later eighteenth century was a copying machine patented by James Watt in 1780. It was a device more accurately known as a copying press, and consisted of a mechanism for exerting pressure on a dampened sheet of tissue placed over the document to be copied. Written with ink based on gum arabic or sugar, the image on the original was transferred to the tissue copy, albeit in reverse, when the two sheets were interfaced under pressure. The reason for the tissue was so that the copy could be viewed correctly looking through it from the back.

A copying press that Benjamin Franklin brought back to the U.S. from Europe was apparently one of this type. George Washington was known to have had two copying presses. Thomas Jefferson, a prolific letter writer, made copies of his correspondence using a copying machine he considered not only "a most precious possession" but "the finest invention of the present age." Jefferson first used a copy press and then a pantograph, or "polygraph" as his was known.

The polygraph was a mechanical apparatus that used wires and movable wooden arms holding a pen or pens to duplicate, on a separate page or pages, the motion of the human arm writing out the original. One of the earliest U.S. patents was for a device of this kind—a "machine for writing with two pens" patented in 1799 by Marc Isambard Brunel.

An advantage to the polygraph was that it made exact (not image-reversed) copies on plain paper, the same paper as the original if desired. But the polygraph was a delicate, fragile mechanism that was difficult and clumsy to use.

It was versions of the copying press, rather than the polygraph, that generally became the standard of nineteenth century. Besides being simpler to use, the copying press could be made small and rugged enough to be easily transportable, and it was sometimes used by travelers.

A device known as the electric pen, patented by Thomas A. Edison in 1876, led to the most common form of copying machine of the early twentieth century. Developed as part of Edison's

automatic telegraph, the electric pen, working at a rate of roughly 8000 pulses per minute, could make minute perforations in the form of letters or drawings on a stencil. A plain sheet of paper was then placed under the stencil, and ink was pressed through with a roller, making exact copies, albeit in small quantity. This led to the mimeograph machine, the mainstay of copying over approximately the first half of the twentieth century. At first, copies were made by hand, one by one. The rotary mimeograph, which was introduced by A.B. Dick in 1904, made copies automatically using a revolving cylinder, at a great increase in speed.

The photostat process, using a special camera to produce an image directly on photosensitized paper without going through a negative, was developed in the early twentieth century. Photostats remained a common way of copying documents until the coming of the modern photocopier. A shortcoming of the photostat, in comparison to the photocopier, is that it was not ordinarily consumer-operated technology.

Do-it-yourself copiers suitable for the small office began to appear in significant numbers by the 1950s. Ordinarily, photosensitized paper was required. Some processes worked with a liquid and others with fumes, and still others with infrared rays to produce heat by which the image was transferred to photosensitized paper (thermography). In general, copies were clearly identifiable as such, the paper usually having a sheen and glossy feel unlike, and inferior to, modern plain-paper copies.

Meanwhile xerography, the principal modern form of photocopying, was also under development. Generally speaking, xerography (from the Greek *xeros* meaning "dry") uses a dry powder as opposed to ink or liquid chemicals. Static electricity attracts and bonds the powder to form an image on paper, and heat then makes the bond permanent.

When xerography came into use, it was generally only for large office applications. The first commercial XeroX Copier (then spelled with a capital X at the end) was introduced in 1949 by the Haloid Company. It was based on a 1938 invention of Chester Carlson. Haloid later became Haloid Xerox and subsequently Xerox Corporation.

The first XeroX was messy and difficult to use. Most of the process was carried out manually, and it often misprinted. By modern standards, it was also a notoriously slow process. At an early demonstration, according to one newspaper account, observers timed the operation at 45 seconds per copy.

In 1960 a vastly improved model, the Xerox 914, was introduced, and with it the modern era of photocopying. The push-button 914 worked automatically and printed on plain paper as opposed to the less-desirable photosensitized paper. However, it weighed 290 kilograms, limiting its use to large-office applications. Its bulky size notwithstanding, the 914 caught on quickly and revolutionized photocopying.

Modern xerographic copiers, produced by a number of manufacturers, are available as desktop models suitable for the home as well as the small office. Many modern copiers reproduce in color as well as black and white, and office models can rival printing presses in speed of operation.

See also **Electronics; Printers**

MERRITT IERLEY

Further Reading

A.B. Dick Company. *The Edison Mimeograph*. Pamphlet, A.B. Dick Company, Chicago, 1889.

A.B. Dick Company. *The Story of Stencil Duplication*. Pamphlet, A.B. Dick Company, Chicago, 1935.

Bedini, S.A. *Thomas Jefferson and His Copying Machines*. University Press of Virginia, Charlottesville, VA, 1984.

Carlson, C. The original duplicator. *Technology Review*, September/October 1998.

Ierley, M. *Wondrous Contrivances: Technology at the Threshold*. Clarkson Potter, New York, 2002.

Rhodes, B. and Streeter, W.W., *Before Photocopying: The Art and History of Mechanical Copying*, 1780–1938. Oak Knoll Press, New Castle, DE, 1999.

Verry, H.R. *Document Copying and Reproduction Processes*. Fountain Press, 1958.

Williams, E.M. *The Physics and Technology of Xerographic Processes*. Wiley, New York, 1984.

Photosensitive Detectors

Sensing radiation from ultraviolet to optical wavelengths and beyond is an important part of many devices. Whether analyzing the emission of radiation, chemical solutions, detecting lidar signals, fiber-optic communication systems, or imaging of medical ionizing radiation, detectors are the final link in any optoelectronic experiment or process.

Detectors fall into two groups: thermal detectors (where radiation is absorbed and the resulting temperature change is used to generate an electrical output) and photon (quantum) detectors. The operation of photon detectors is based on the photoelectric effect, in which the radiation is absorbed within a metal or semiconductor by direct interaction with electrons, which are excited to a higher energy level. Under the effect of an electric field these carriers move and produce a

measurable electric current. The photon detectors show a selective wavelength-dependent response per unit incident radiation power.

The photoeffect takes two forms: external and internal. The former involves photoelectric emission where the photogenerated electrons escape from the material as free electrons. In the latter process, the excited carriers remain within the material, usually in semiconductors, thereby increasing its conductivity (photconductive detectors or photoresistors).

The principle of photoemission was first demonstrated in 1887 by Heinrich Hertz who noticed that the breakdown voltage of spark gaps decreased when ultraviolet light was shone on the metal cathode. Two years later, Julius Elster and Hans Geitel revealed that if an alkali metal electrode was used, this effect could be produced with visible radiation. Although these effects could be demonstrated reproducibly, no satisfactory explanation was offered until Einstein proposed his theory of photoemission in 1905.

The first internal photon effect, the photoconductive effect, was discovered by Willoughby Smith in 1873 when he experimented with selenium as an insulator for submarine cables. The material Tl_2S (thallium sulfide) was the first infrared photoconductor of high responsivity and was developed by Theodore W. Case in 1917 as an infrared sensor for signaling. The years during World War II saw the origin of modern detector technology because of the need for signaling and aircraft detection.

Depending on the nature of the interaction, photon detectors can be further subdivided into different types as shown in Table 1. The most important are intrinsic detectors (in which an electron or hole moves from the valence band to the conduction band), extrinsic detectors (in which transitions are between the doped level in energy gap and conduction or valence bands), photoemissive detectors (like PtSi Schottky barrier detectors), and quantum well detectors (with intersub-band transitions inside bands).

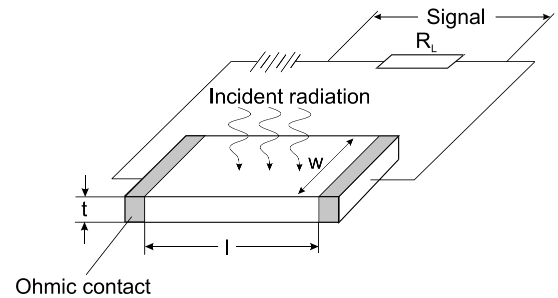


Figure 3. Geometry and bias of a photoconductive detector.

The light-generated carriers can produce an electric signal in two ways. In one case they can increase the sample conductivity. Such devices are called photoconductive detectors or photoresistors (Figure 3). The other way to produce a signal is to use a semiconductor device that has an internal electric field. It may be a *p–n* junction diode (Figure 4), or a Schottky barrier (metal semiconductor contact) diode (Figure 5). In any of these cases, electrons and holes can be separated by the built-in electric field or, more typically, by the electric field that is a combination of the external and internal electric fields. Such a photovoltaic detector is called a photodiode.

In comparison with photoconductive detectors, photodiodes exhibit four important advantages:

1. Low or zero bias voltage
2. High impedance, which aids coupling to read-out circuits in imaging arrays
3. Capability for high-frequency operation

Table 1 Photon detectors with internal photoeffect.

Type	Transition	Electrical output	Example
Intrinsic	Interband	Photoconductive Photovoltaic Capacitance (MIS)	Si, GaAs, GaN, PbSe, InSb, HgCdTe Si, GaN, InGaAs, InSb, HgCdTe Si, GaAs, InSb, HgCdTe
Extrinsic	Impurity to band	Photoconductive	Si:As, Si:Ga, Ge:Cu, Ge:Hg
Free carriers	Intraband	Photoemissive	PtSi, Pt_2Si, IrSi Schottky barriers GaAs/CsO
Quantum wells	To or from spatially quantised levels	Photoconductive Photovoltaic	GaAs/GaAlAs, InAs/InGaSb

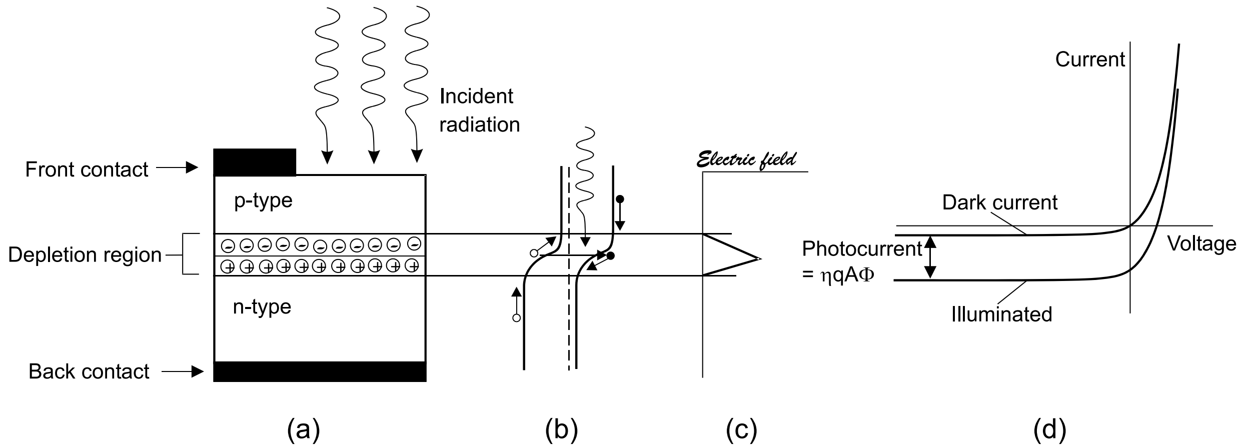


Figure 4. *p–n* junction photodiode: (a) structure of abrupt junction, (b) energy band diagram, (c) electric field, and (d) current–voltage characteristics.

4. The compatibility of the fabrication technology with planar-processing techniques.

Photoemissive detectors are generally the detectors of choice for ultraviolet (UV), visible, and near-infrared applications where high efficiency is available. They usually take the form of vacuum tubes called phototubes. Electrons are emitted from the surface of a cathode and travel to an electrode (anode), which is maintained at a higher electric potential (Figure 6(a)). As a result of the electron transport between the cathode and anode, an electric current proportional to the photon flux incident on the photocathode is created in the circuit. The photoemitted electrons may also impact other specially placed metal or semiconductor surfaces in the tube, called dynodes, from which a cascade of electrons is emitted by the process of secondary emission. The result is an amplification of the generated electric current by a factor as high as 10^7. This device is known as a photomultiplier tube (Figure 6(b)).

The secondary emission is also used in a modern imaging device called the microchannel plate, widely used to detect UV radiation, and soft x-ray fluxes. It consists of a honeycomb array of millions of glass capillaries with an internal diameter of around 10 micrometers drawn out by fiber optic techniques, in a glass plate of a thickness of around 1 millimeter. Both faces of the plate are coated with thin metal films that act as electrodes and a voltage is then applied across them (Figure 6(c)). The interior walls of each capillary are coated with a secondary-electron-emissive material and behave as a continuous dynode, multiplying the photoelectron current emitted at that position (Figure 6(d)). In such a way, the local photon flux can be converted into a substantial electron flux that can be measured directly. Furthermore, the electron flux can be reconverted into an optical image by using a phosphor coating as the rear electrode to provide electroluminescence; this combination provides an image intensifier.

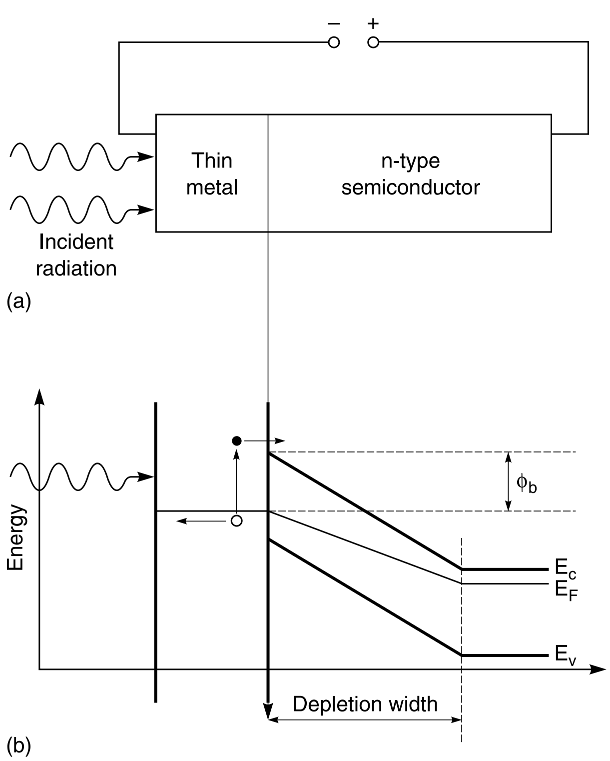


Figure 5. Schottky barrier photodiode: (a) structure and (b) energy band diagram.

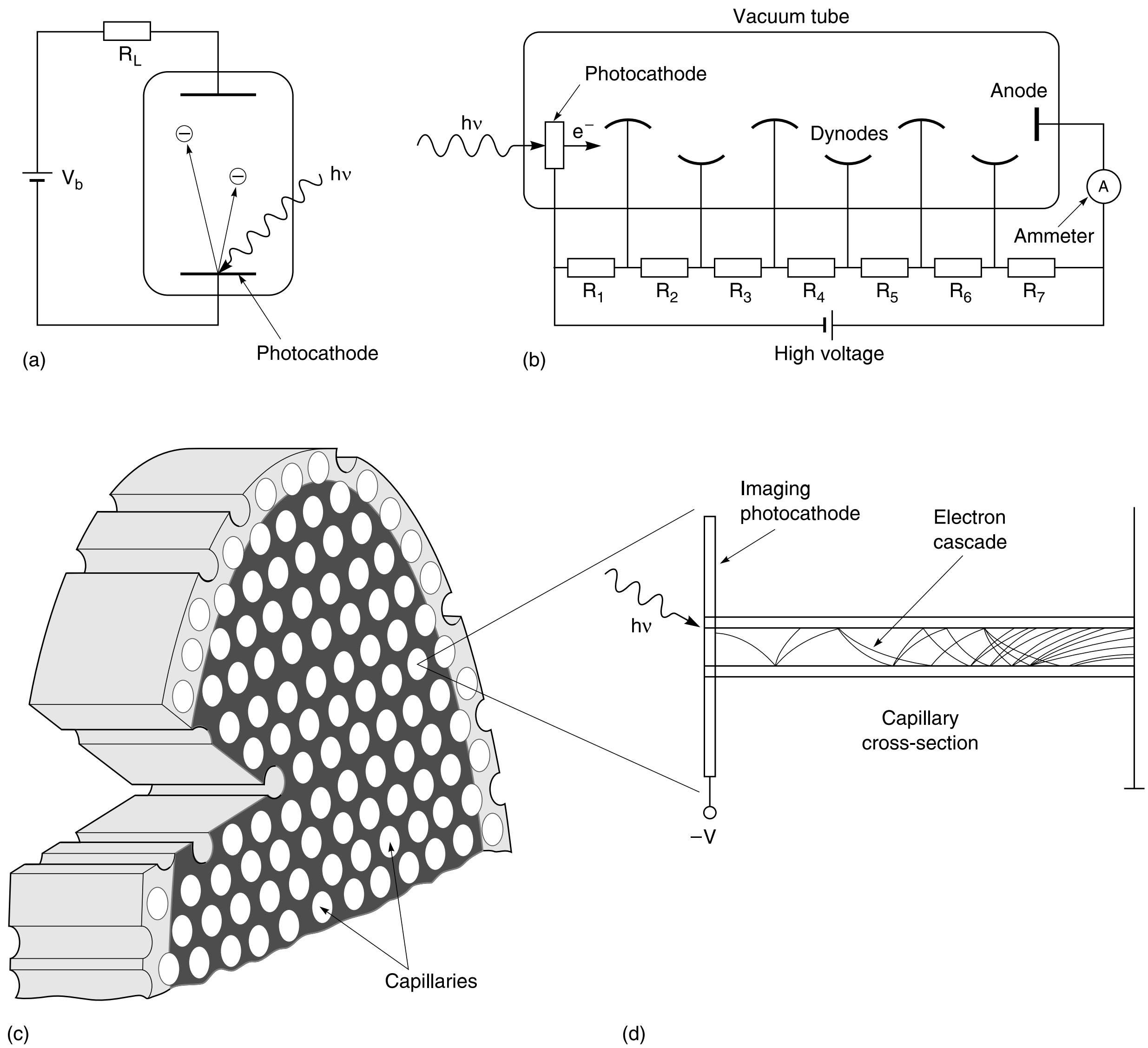


Figure 6. Schematic presentation of: (a) phototube, (b) photomultiplier tube, (c) cutaway view of microchannel plate, and (d) single capillary in a microchannel plate.

Photoconductive detectors and especially photodiodes (with their very low power dissipation) are combined with electronic readouts to make detector arrays used in imaging devices. They are often called as focal plane arrays (FPAs) because they are located in the focal plane of imager optics.

There are a number of architectures used in the development of FPAs. In general, they may be classified as monolithic or hybrids. In the monolithic approach, both detection of light and signal readout (multiplexing) is done in the detector material. The integration of detector and readout onto a single monolithic piece reduces the number of processing steps, increases yields, and reduces costs. For visible and near-infrared detection, monolithic structures can be built in silicon, forming arrays with more than 10 million high-performance pixels. The most highly developed of these visible detectors is the charge-coupled device (CCD). This approach to image acquisition was first proposed in 1970 by Bell Lab researchers Willard S. Boyle and George E. Smith.

In effect, a CCD is a light-sensitive device that stores electrical signals. The CCD technique relies on the optoelectronic properties of a well-established semiconductor architecture: the metal oxide semiconductor (MOS) silicon capacitor (Figure 7). An MOS capacitor typically consists of an extrinsic silicon substrate (in this case, *p*-type) on which is grown an insulating layer of silicon dioxide (SiO_2).

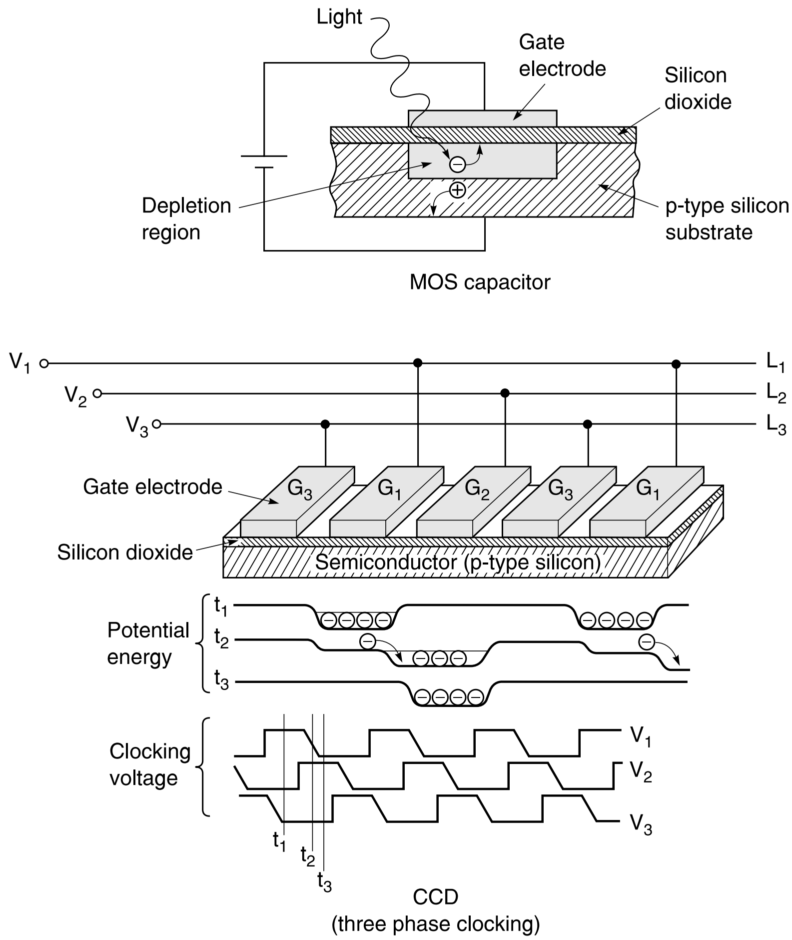


Figure 7. Charge-coupled device. Light absorbed by a *p*-type metal oxide semiconductor (MOS) silicon capacitor (top) creates mobile electrons that become trapped in the depletion region under the gate. Three-phase CCDs (bottom) use a timed sequence of three gate voltages to move accumulated electrons across an array of MOS capacitors. Note the potential energies at times t_1, t_2, and t_3. Gates G_1, G_2, and G_3 define a single pixel.

When a bias voltage is applied across MOS structure, majority charge carriers (holes) are pushed away from the Si–SiO_2 interface directly below the gate, leaving a region depleted of positive charge and available as a potential energy well for any mobile minority charge carriers (electrons). Electrons generated in the silicon through absorption will collect in the potential energy well under the gate. Linear or two-dimensional arrays of these MOS capacitors can therefore store images in the form of trapped charge carriers beneath the gates. In the CCD solution the accumulated charges are transferred from potential well to the next well by using sequentially shifted voltage on each gate. One of the most successful voltage-shifting schemes is called three-phase clocking (Figure 7, bottom). Column gates are connected to the separate voltage lines (L_1, L_2, L_3) in contiguous groups of three (G_1, G_2, G_3). The setup enables each gate voltage to be separately controlled.

The first CCD imager sensors were developed in the 1970s primarily for television analog image acquisition, transmission, and display, and CCD cameras were first used in telescopes in 1979. CCD cameras have subsequently been developed for scanners, video cameras, and digital cameras. Modern CCD detectors have excellent x-ray response, and have been used in orbiting x-ray telescopes. With increasing demand for digital image data, the traditional analog raster scan output of image sensors is of limited use, and there is a strong motivation to fully integrate the control, digital interface, and image sensor on a single chip. In particular, silicon fabrication advances for computer processors and memory now permit the implementation of complementary MOS (CMOS) transistor structures that are considerably smaller than the wavelength of visible light and have enabled the practical integration of multiple transistors within a single image sensor. At each CMOS imager sensor pixel, there is a photosensor with some integrated MOS transistors that are used to multiplex and possibly manipulate the analog signal from the photosite (Figure 8). Analog output signals from any CMOS imager

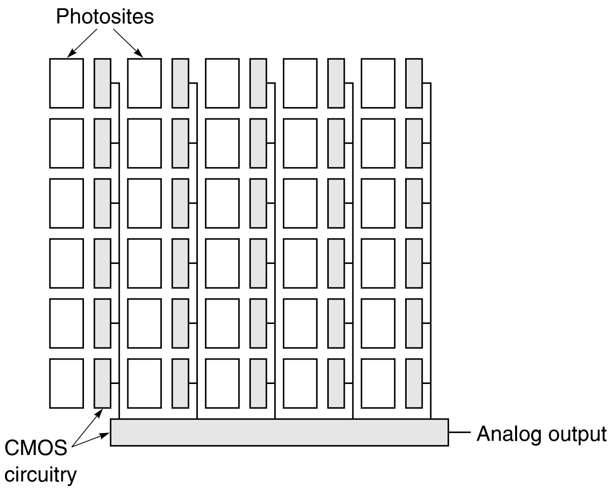


Figure 8. CMOS imager architecture.

pixel can be randomly accessed through wires and multiplexing circuitry.

Developing rapidly, CMOS imaging technology has several potential advantages over CCD technology: low-voltage operation and low power consumption, compatibility with integrated CMOS circuitry, random access to image data, and a lower cost for both digital video and still camera applications. The processing technology for CMOS image sensors is typically two to three times less costly than standard CCD technology.

Both CCD and CMOS imagers described above are monolithic devices. In infrared spectral range, hybrid architectures are usually used. Hybrid FPA detectors and multiplexers are fabricated on different substrates and mated with each other by the flip-chip bonding or loophole interconnection. In this case we can optimize the detector material and multiplexer independently. Other advantages of the hybrid FPAs are near 100 percent fill factor and increased signal-processing area on the multiplexer chip. In the flip-chip bonding, the detector array is typically connected by the contacts via indium bumps to the silicon multiplex pads.

See also **Iconoscope; Infrared Detectors; Semiconductors, Elemental; Semiconductors, Pre-Band Theory; Solar Power Generation**

ANTONI ROGALSKI

Further Reading

Bhattacharya, P. *Semiconductor Optoelectronic Devices*. New Jersey, Prentice Hall, 1997.

Csorba, I.P. *Image Tubes*. Sams, Indianapolis, 1985.

Donati S. Photodetectors: Devices, Circuits and Applications. Prentice Hall, New Jersey, 2000.

Goldstein, A. Charge coupled device, in *Instruments of Science*, Bud, R., Ed. Garland, London, 1998.

Levine, B.F. Quantum-well infrared photodetectors. *J. Appl. Phys*. 74, R1–R81, 1993.

Rieke, G.H. *Detection of Light: From the Ultraviolet to the Submillimeter*. Cambridge University Press, Cambridge, 1994.

Rogalski, A. *Infrared Detectors*. Gordon & Breach, Amsterdam, 2000.

Sequin, C. and Tompsett, M. *Charge Transfer Devices*. Academic Press, New York, 1975.

Smith, G. and Boyle, W. The invention and early history of the CCD, in *Optical Detectors for Astronomy*, Amico, P. and Beletic, J.W. Kluwer, Dordrecht, 1998.

Plastics, Thermoplastics

Thermoplastic polymers are perhaps some of the best-known plastics, as they are used for a variety of household items and packaging materials. They comprised 10 percent of the global chemical industry in 1995 (around 90 million tons). Initially made from a process based on coal, they are now usually made from oil-based products.

Thermoplastics include plastics developed as early as 1877 when polymethylmethacrylate (PMMA, or acrylic) was first formed by Rudolph Fittig, developed by Otto Rohm, a German chemist, but not commercially developed by Rohm & Haas until 1928 and later by Rowland Hill and John Crawford at ICI in 1934. In the 1930s other thermoplastics were developed, such as polyethylene by ICI in the U.K., and nylon 6 and nylon 6,6 by WalIace Hume Carothers at DuPont in the U.S. Polyethylene is now commonly used as a packaging material and synonymous with the plastic bag. Nylon was first used to make toothbrush bristles in 1938, then ladies' stockings in 1939. Thermoplastic polypropylene, polyvinyl chloride (PVC), and polystyrene are common packaging materials today. Carothers had worked on polyesters and his work was progressed by John R Whinfield and James T Dickson at the Calico Printers Association in the U.K., who produced the polyester fiber Terylene in 1941. This material is widely used to make clothing and soda bottles. Thermoplastics also include certain resins such as polyether ether ketones (PEEK) which are used to make composite material with superior mechanical properties although at a high cost, therefore they are normally only used in high-technology fields such as aerospace (see Composite Materials).

With thermoplastics, increased strength and thermal resistance comes at an increased price, with polystyrene as the cheapest, least thermal resistant and weakest thermoplastic, followed by low-density polyethylene, high-density polyethy-

lene, polypropylene, polymethyl methacryalate (PMMA, or acrylic), high-impact polystyrene, acrylonitrile butadiene styrene (ABS*) (used for car bumpers), polyesters, polycarbonates and polysulfones to nylon (a polyamide) which has the highest strength and thermal resistance but is also the most expensive. Polycarbonates have high impact strength, hardness, toughness, and resistance to temperatures between about −40°C and 145°C. Polysulfones are heat-resistant at temperatures of up to 150°C

Thermoplastics can be molded to set into a certain shape, but they can then be reshaped after reheating. This is due to their molecular structure, which consists of long chains of molecules held together by weak intermolecular forces. They have a structure almost resembling spaghetti, bound together by the weak Van der Waals forces. Some cross-linking can occur with side groups such as the vinyl group in PVC. The orientation of these side groups will have a great influence on how the polymer will behave and its properties such as strength.

Thermoplastics have to be processed at a higher temperature than thermosetting plastics; this may be more than 400°C for "high temperature" thermoplastics.

The weak intermolecular forces in thermoplastics make them relatively easy to process using a variety of methods ranging from extrusion, vacuum and blow molding to perhaps the most common process, injection molding. They are pliable, and easily shaped and molded. The molding process becomes easier as thermoplastics become hotter, but at some point they will melt.

Thermoplastics have a glass transition temperature (Tg) above room temperature, whereas that of elastomers is below room temperature. The glass transition temperature is where a polymer changes from a rigid solid to a rubber. Below this temperature the polymer becomes hard and brittle, like glass. Some polymers are used above their glass transition temperatures while others such as polystyrene and PMMA are used below.

Thermoplastics can also be blended with other polymers, such as elastomers, to enhance certain properties such as toughness. These are known as copolymers and they can be engineered for specific purposes.

There is also a class of thermoplastic elastomers with physical rather than chemical cross-links. An example of such a material is a thermoplastic elastomer, polyurethane.

Thermoplastics are easy to mold and can be shaped. They are resistant to deformation but will eventually permanently deform or break. They are hard and brittle below their Tg, but pliable and soft above it. Crystalline polymers melt below their glass transition temperature, whereas amorphous polymers are hard and brittle below their Tg, but become flexible and rubbery above it. Most thermoplastics are a mix of a crystalline and an amorphous structure. Hard plastics such as polystyrene and PVC are used at temperatures below their Tg whereas flexible plastics such as polypropylene are used above it. Some thermoplastics such as polypropylene, nylon, polyketones and syndiotactic polystyrene are highly crystalline, whereas others such as PMMA, polycarbonates and atactic polystyrene are highly amorphous due to their polymer structure and the intermolecular forces.

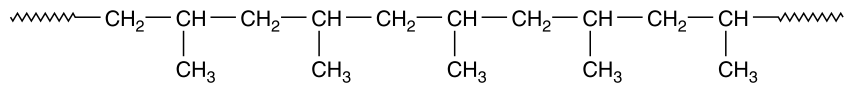
Thermal or mechanical methods can be used to alter the crystalline structure of a thermoplastic. If a polymer is cooled slowly from its melting point, a higher degree of crystallinity is more probable. However when polymers are cooled quickly from the melt, the amorphous chains may be frozen into the solid. This is because the chains will not have had enough time to untangle from their melted form to separate and form crystals.

Thermoplastic composites are frequently tougher than thermoset composites, but although they do not possess enhanced fatigue or static properties and they may also have worse compression strength, they are more resistant to moisture and a range of industrial solvents than thermosets.

Thermoplastics are made from molecular chains with various types of stereochemical arrangements possible that can confer different properties on the polymer. These are called atactic, isotactic and syndiotactic. In an isotactic polymer, all the substituent groups lie on one side of the molecular chain. In isotactic polypropylene, for example, the methyl (CH_3) groups are on the same side of the molecular chain. This arrangement allows the molecular chains to pack together more easily giving a crystalline structure that is strong, stiff and brittle (see Figure 9).

Atactic thermoplastics have substituent groups that are randomly placed on both sides of the chain for example the methyl groups in atactic polypropylene (see Figure 10). This arrangement means that the chains are unable to pack together, resulting in an amorphous structure that makes the polymer tough and rubbery. This results in a form of toughened polypropylene.

Syndiotactic polymers have repeating units on either side of the molecular chain backbone,



Isotactic polypropylene

Figure 9. Chemical formula. [*after http://www.psrc.usm.edu/macrog/pp.htm*]

alternating with each other (see Figure 11). In syndiotactic polystyrene, the chains are able to pack tightly together giving a crystalline structure whereas the atactic form of polystyrene is amorphous as the chains cannot pack together so tightly.

Thermoplastics can be recycled, but require careful sorting to separate the various different polymers. If different grades are mixed the resultant recycled polymer will have variable properties. Normally, recycling a polymer will involve some deterioration in its properties.

Thermoplastics have flexibility and so are useful where this is needed, for example in squeezable washing-up bottles. The first of these made was the 1958 "Squeezy" bottle, which was made of polyethylene (although with metal ends at this point). Thermoplastics can be blow-molded into a variety of shapes and polyethylene terepthalate (PET) is a popular thermoplastic used to make "pop" bottles.

Thermoplastics are excellent materials for coextrusion—a process widely used in packaging to make multilayered sheet with different properties in the different layers—tough on the outside and impermeable on the inner layers. This technique is popular for producing packaging materials.

Although polyethylene is self-toughening, different formulations of polyethylene have been developed such as low-density polyethylene, the earliest type developed first in 1933, high-density polyethylene, and now ultrahigh molecular weight polyethylene.

Polypropylene is used for packaging and in cars and polypropylene fibers are also used for clothing, carpets and nonwoven fabrics. Impact-resistant copolymers of polypropylene are used to make car bumpers as well as for medical use.

Teflon (polytetrafluoroethylene or PTFE) is used to make products that operate at high temperatures because of its heat-resistant properties. It has excellent chemical, electrical, mechanical and thermal properties and is also chemically inert. Almost nothing sticks to Teflon hence its use for non-stick frying pans. As it is very heat resistant, it is often used in space applications, for example as a material to protect the deployment rods of the solar arrays which replaced the original arrays during the first servicing mission of the Hubble space telescope in 1993.

The future of thermoplastics demands improved recyclability, toughness, repairability and "smart applications" such as self-healing abilities and perhaps inbuilt obsolescence and, ideally, biodegradability. Bioengineered thermoplastics—the class of so-called biopolymers—are an exciting new development. Polyesters are already engineered to make specialist fibers required for sporting applications. More improved engineered fibers are continuously being developed. In the future they will also have to possess an environmentally friendly life-cycle.

See also **Biopolymers; Fibers, Synthetic and Semi-Synthetic; Plastics, Thermosetting; Synthetic Resins**

SUSAN MOSSMAN

Further Reading

Arlie, J.-P. *Commodity Thermoplastics*. Editions Technip, Paris, 1990.

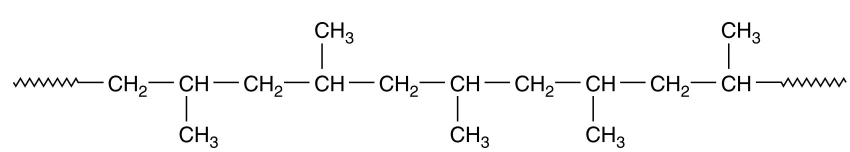
MacDermott, C.P. and Shenoy, A.V. *Selecting Thermoplastics for Engineering Applications*, 2nd edn, vol. 42. 1997.

Olabisi, O. *Handbook of Thermoplastics*, vol. 41. 1997.

Scheirs, J. *Polymer Recycling*. Wiley, Chichester, 1998.

Plastics, Thermosetting

Thermosetting plastics, or thermosets are a type of polymer, usually made by mixing two or more suitable liquids together. Thermosets are usually supplied in the form of partly polymerized pre-



Atactic polypropylene

Figure 10. Chemical formula. [*after http://www.psrc.usm.edu/macrog/pp.htm*]

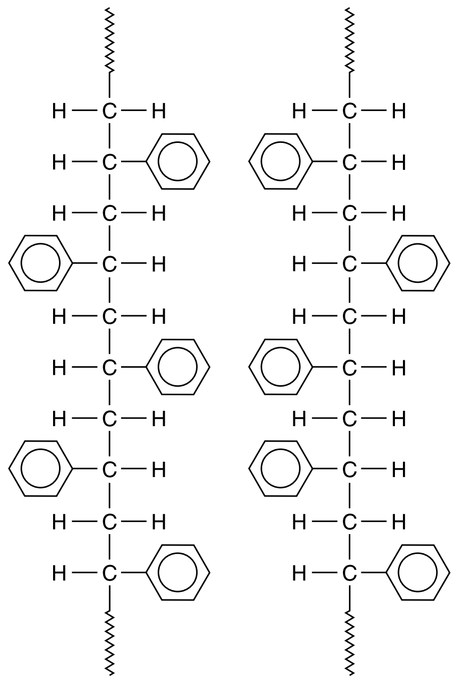


Figure 11. Chemical formula.
[*after http://www.psrc.usm.edu/ macrog/styrene.htm*]

cursors or as mixtures of monomer and polymer. When mixed, the liquids undergo a chemical reaction and form a hard solid. Polymers are made up of long molecules that resemble chains. In thermosetting polymers, the long molecular polymer chains are linked by covalent bonds located in three dimensions by cross-linking. When the liquid precursors are mixed together, bonds start to form between the polymer chains during the curing process. This process continues until all the chains have joined together forming a single giant molecule. The molecule's gigantic size makes it solid. Chemicals, applied heat or radiation may be used during manufacture to bring about the polymerization process (curing).

Thermosets have a critical temperature and once heated above this temperature and molded into shape, they will stay in that form. Usually, thermosets cannot be altered by further heating. However, although the two-step curing process forms a three-dimensional structure with cross-linked bonds that do not break down on heating, at very high temperatures thermosets will break down permanently.

The molecular structure of thermosetting polymers determines their properties. The cross-links that exist in their molecular structure stop the molecular chains from sliding past one another, and give thermosets a higher modulus and better creep resistance. Once it has begun, the cross-linking process cannot be reversed. This results in materials that cannot be recycled by remelting. Thermosets are usually more brittle, less flexible and impact-resistant than thermoplastics although they possess better abrasion and dimensional properties.

Thermosets are similar to elastomers—at room temperature, the polymer chains in thermosets are below their glass transition temperature (Tg), making them hard and brittle. However, the polymer chains in elastomers are above their Tg at room temperature, and this factor makes them rubbery. The Tg is the point at which a polymer changes from a rigid solid to a rubber. Some polymers are used above their Tg while others are used below it.

Thermosets in the form of phenolic resins were first developed in the U.S. by the Belgian émigré chemist, Leo Baekeland in 1907 and first patented in 1909. They were developed contemporaneously in Britain by Sir James Swinburne. Baekeland named his invention Bakelite. He reacted phenol and formaldehyde under controlled conditions, producing an amber-colored resin to which he added a filler such as wood flour or cotton flock, producing dark-colored moldings: normally dark brown, dark red or dark green. Products made from phenolic resins possess good electrical resistance and mechanical properties, and Bakelite was immediately utilized as an insulating material in electrical applications such as plugs and insulators. As phenolics are very difficult to ignite and are thermosets, they were also used for purposes that require materials that need to be able to resist high temperatures, such as in thermos flasks.

Other thermosetting polymers developed since the advent of phenolic resins in the early twentieth century include amino resins (including thiourea-urea-, urea- and melamine-formaldehydes), unsaturated polyesters, polyurethanes, and epoxy

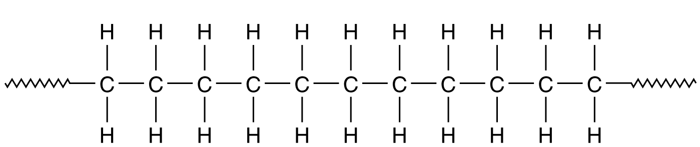


Figure 12. The polyethylene molecule.

resins. Thiourea-urea formaldehydes were discovered in 1924 by Edmund Rossiter, while working for the British Cyanides Company and became popular for a range of decorative tableware called Bandalasta. By 1929, urea-formaldehyde had been developed with better properties. Baron Justus von Liebig, a German chemist, discovered melamine resin in 1834, but melamine-formaldehyde polymer was not patented until 1935. The American Cyanamid Company produced it commercially in 1939. As melamine-formaldehyde was water resistant and tougher than urea-formaldehyde and transparent, it became possible to impregnate patterned papers for surfacing decorative laminates such as Formica and Warerite. These led the way to easy-to-care for surfaces, particularly in the kitchen. Polyimides were introduced in the 1960s and cyanate esters are now under development (see Figure 13).

As thermosets are cross-linked polymers that do not melt, they have to be molded under pressure; for example, by compression molding, transfer molding and various injection molding techniques. Thermoset composite materials can be made in a number of ways, for example by using processes such as resin transfer molding, centrifugal molding and autoclave molding. The manufacture of thermoset composite materials may be rather labor intensive in cases where some components have to be hand laid. For example when making a carbon fiber–thermosetting resin composite, the carbon fiber might be laid by hand, usually in the form of a prewoven fiber. The thermosetting resin, such as epoxy, will then have to be applied and the system allowed to cure. Another type of thermoset composite material such as glass-reinforced fiber (GRP) might use short fibers that are embedded in a polyester or epoxy resin matrix.

The chemistry of processing of thermosets is more complicated than that needed to process thermoplastics, which need only to be melted and cooled. As thermosets are transformed from materials that can be fused and are soluble to highly inflexible cross-linked resins that are impossible to mold, they have to be manufactured during the cross-linking process.

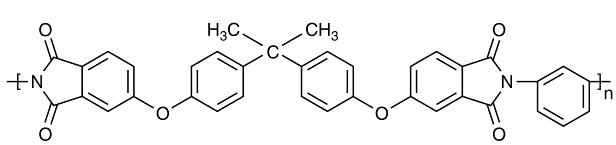


Figure 13. A polyimide molecule
[*after http://www.psrc.usm.edu/macrog/imide.htm*]

Unsaturated polyester resins have a maximum service temperature of around 100°C and vinyl esters and epoxies a temperature of about 150°C. For this reason, high-performance systems such pyromelltimide-type polyimides, cyanate esters and *bis*-maleimides have been developed, as resin matrices that can be used at higher temperatures are needed in certain applications.

These high-performance systems form thermosets that can be used at temperatures of up to 250°C and molded at around 250–300°C. In contrast, high-performance thermoplastics have to be processed at higher temperatures of up to 450°C.

Jon Jakob Berzelius produced the first polyester resin (polyglycerol tartrate) in 1847. A range of polyester resins is now available, including the Marco and Crystic unsaturated polyester resins, which were developed by Scott Bader in 1946. The first commercial use of low-pressure resins to make reinforced polymer composites occurred in 1942 in the form of glass-cloth-reinforced resin radomes made for aircraft in the U.S. By the late 1940s GRP, more commonly known as fiber glass, was used commercially, with many of the earliest developments derived from making hulls. The first car with a fiber-glass body, the Corvette, was made in 1953. Fiber glass was also used for corrugated roofing, decorative moldings and (unsuccessfully) for window frames and baths. Certain items of furniture, such as stackable chairs, are often made using fiber glass and first appeared in the 1950s.

Epoxy resins are also used with glass fiber to make a composite low-pressure reinforced molding material. Epoxy resins were first developed in the 1930s by Pierre Castan and became commercially viable in 1939 with IG Farben's patent concerning liquid polyepoxides. The initially high cost of production of epoxy resins as compared with polyesters limited their use until later improvements in production methods. Today they are particularly advantageous for space applications, due to their light weight and excellent electrical properties.

Thermosets are particularly suitable as materials of choice for making big components such as boat hulls, as the liquid precursors from which they are made are not very viscous and so can flow easily, filling up large moulds without needing great injection pressures. Once they have filled up the mould, the precursors then react to form a solid product.

Thermosets such as phenolics are very difficult to recycle because of their high temperature

resistance, inability to soften on melting, and their insolubility. However Bakelite (phenol formaldehyde) will biodegrade if broken, provided it has been filled with organic fillers such as woodchip, which can become the site for biodegradation to occur. Thermoset composites are used in a variety of products from GRP furniture, boat hulls, and sheet-molded compounds (SMC) to carbon fiber-reinforced plastic (CFRP) tennis rackets, Formula One car bodies and aircraft radomes. These are currently difficult to recycle. Furniture and sports equipment may be recycled in the form of repair or reuse, but eventually, once beyond repair, they end up on the scrapheap for the foreseeable future. Thermoset composite materials can be burned as fuel or to give energy (pyrolysis). However, because many of these fiber-based products such as CFRPs are so expensive to produce, ways have been found to recycle them by grinding and reusing them as fillers, and by selective chemical degradation.

See also **Plastics, Thermoplastics; Synthetic Resins**

SUSAN MOSSMAN

Further Reading

CES. *Cambridge Engineering Selector*. Granta Design, Cambridge, 2000.

Goodman, S.H., Ed. *Handbook of Thermoset Plastics*. Noyes Publications, Westwood NJ, 1998.

Mayer, R.M. *Design with Reinforced Plastics*. Design Council, London, 1993.

Morton-Jones, D.H. *Polymer Processing*. Chapman & Hall, London, 1989.

Scheirs, J. *Polymer Recycling*. Wiley, Chichester, 1998.

Positron Emission Tomography

Medical imaging is the term for a number of techniques for viewing anatomical structures and, by the end of the twentieth century, physiological functioning of the body. Medical techniques such as x-ray, computed tomography, and magnetic resonance imaging yield exquisitely detailed images. Such images can be acquired by viewing the decay of radioisotope bound to molecules with known biological properties. This class of imaging techniques is known as nuclear medicine imaging or emission computed tomography (ECT). In ECT multiple cross-sectional images of tissue function can be produced, thus removing the effect of overlying and underlying radioactivity. Experimentation with x-ray tomography began as early as the 1930s, but the technology was not commonly used in clinical medicine until the 1980s.

The techniques of ECT are usually considered as two separate modalities.

1. Single photon emission computed tomography (SPECT), invented in the 1970s, involves the use of a radioisotope such as technitium-99 ($^{99}Tc^{m}$), where a single γ-ray is emitted per nuclear disintegration.
2. Positron emission tomography (PET), the development of which began in the 1950s. PET is a radiotracer imaging technique, in which tracer compounds labeled with positron-emitting radionuclides are injected into the subject of the study. These tracer compounds can then be used to track biochemical and physiological processes *in vivo*.

One of the prime reasons for the importance of PET in medical research and practice is the existence of positron-emitting isotopes of elements such as carbon, nitrogen, oxygen and fluorine, which may be processed to create a range of tracer compounds that are similar to naturally occurring substances in the body. Among many applications, these radiotracers, such as carbon-11 (^{11}C), nitrogen 13 (^{13}N), oxygen 15 (^{15}O), and fluorine-18 (^{18}F), are used clinically to identify and diagnose cancers, epilepsy and other movement disorders, and heart and cardiovascular problems.

After the injection of a tracer compound labeled with a positron-emitting radionuclide, the subject of a PET study is placed within the field of view (FOV) of a number of detectors in a scanner that are capable of registering incident gamma rays. The radionuclide in the radiotracer decays, and the resulting positrons subsequently annihilate on contact with electrons after traveling a short distance (around 1 millimeter) within the body. Each annihilation produces two 511-kiloelectron-volt photons traveling in opposite directions, and these photons maybe detected by the detectors surrounding the subject. The detector electronics are linked so that two detection events occurring within a certain time window may be called coincident and thus be determined to have come from the same annihilation (Figure 14). These coincident events can be stored in arrays corresponding to projections through the patient and reconstructed using standard tomographic techniques. The resulting images show the tracer distribution throughout the body of the subject. Photon isotopes may decay via positron emission, in which a proton in the nucleus decays to a

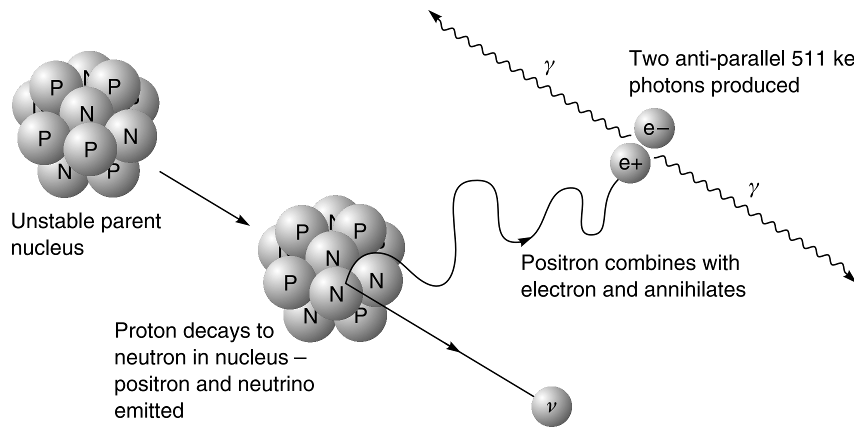


Figure 14. Positron emission and annihilation.

neutron, a positron, and a neutrino. The daughter isotope has an atomic number one less than the parent.

In the PET medical technique, the subject is surrounded by a cylindrical ring of detectors with a diameter of 80 to 100 centimeters and an axial extent of 10 to 20 centimeters. The detectors are shielded from radiation from outside the field of view by relatively thick lead end-shields. Most scanners can be operated in either a slice collimated mode, where axial collimator is provided by thin annual rings of tungsten called septa or in a fully three-dimensional mode where the septa are retracted and coincidences can be collected between all possible detector pairs. The usual configuration of detectors in PET is a rectangular bundle of crystals, termed a block, optimally coupled to several photomultiplier tubes (PMTs). When a photon interacts in the crystal, electrons are moved from the valence band to the conduction band. These electrons return to the valence band as impurities in the crystal, emitting light in the process. Since the impurities usually have metastable excited states, the light output decays exponentially at a rate characteristic of the crystal. The ideal crystal will have a high density so that a large fraction of incident photons scintillate, high light output for positioning accuracy, fast rise time for accurate timing, and a short decay time so that high counting rate can be handled. Most current scanners use bismuth-germanate (BGO), which generates approximately 2500 light photons per 511-kiloelectron-volt photon and has a decay time of 300 nanoseconds. One such block couples a seven by eight array of BGO crystals to four PMT where each crystal is 3.3 millimeters wide in the transverse plane, 6.25 millimeters wide in the axial dimension, and 30 millimeters deep. The block is fabricated in such a way that the amount of light collected by each PMT varies uniquely depending on the crystal in which the scintillation occurred. Hence integrals of the PMT outputs can be decoded to yield the position of each scintillation. The sum of the integrated PMT outputs is proportional to the energy deposited in the crystal.

Coincident events in PET fall into four categories: true, scattered, random, and multiple (Figure 15):

- True coincidence: occurs when both photons from an annihilation event are detected by detectors in coincidence, neither photon undergoes any form of interaction prior to detection, and no other event is detected within the coincidence time window.
- Scatter coincidence: occurs when at least one of the detected photons undergoes at least one scattering event prior to detection. Scatter coincidences add a background to the true coincidence distribution that changes slowly with position, decreasing contrast, and causing the isotope concentrations to be overestimated. The number of scattered events detected depends on the volume and attenuation characteristics of the object being imaged and on the geometry of the camera.
- Random coincidence: occurs when two photons not arising from the same annihilation event are incident on the detectors within the coincidence time window of the system. Also, the number of random coincidence depends on the volume and attenuation characteristics of the object being imaged and on the geometry of the camera.
- Multiple coincidence: occurs when more than two photons are detected in different detectors within the coincidence resolving time.

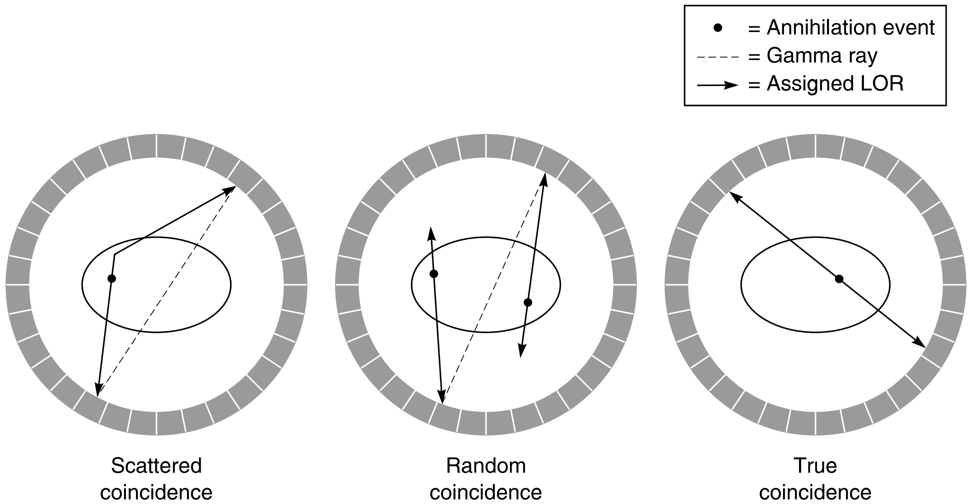


Figure 15. Types of coincidences in positron emission tomography.

See also **Neurology; Nuclear Magnetic Resonance (NMR, MRI); Tomography in Medicine**

STELIOS ZIMERAS

Further Reading

Bailey, D.R. and Meikle, S.R. A convolution–subtraction scatter correction method for 3D PET. *Phys. Med. Biol.*, 39, 411–424, 1994.

Raylman, R.A., Hammer, B.E. and Christensen, N.L. Combined MRI-PET scanner: a Monte Carlo evaluation of improvements in PET resolution due to the effects of a static homogeneous magnetic field. *IEEE Trans. Nuc. Sci.*, 43, 4, 2406–2412, 1992.

Webb, S. *The Physics of Medical Imaging*. Institute of Physics Publishing, Bristol, 1996.

Webb, S. *From the Watching of Shadows: The Origins of Radiological Tomography*. Institute of Physics Publishing, Bristol, 1990.

Power Generation, Recycling

Recovering energy from wastes from municipal or industrial sources can turn the problem of waste disposal into an opportunity for generating income from heat and power sales. The safe and cost-effective disposal of these wastes is becoming increasingly important worldwide, especially with the demand for higher environmental standards of waste disposal and the pressure on municipalities to minimize the quantities of waste generated that must be disposed.

Energy from Waste (EfW) Technologies

There are a number of ways of recovering energy from waste:

- Landfill gas: collecting and using some of the gas that is produced as waste decay in landfill sites. The landfill gas can be used to run engines, fire boilers or kilns, and generate electricity.
- Anaerobic digestion: capturing the gas produced when organic wastes such as food preparation wastes and garden prunings break down in an air-free (i.e. anaerobic) environment. Anaerobic digestion improves on gas collection from landfill sites by using a sealed system where all of the gas can be collected for use as a fuel.
- Refuse-derived fuel: turning the combustible portion of waste, such as paper and plastics, into a fuel that can be stored and transported as pellets or briquettes or directly used on site to produce heat and power.
- Incineration with energy recovery: burning mixed waste in sophisticated plants where the heat given off is harnessed for hot water and electricity generation. Combined heat and power (CHP), which not only produces electricity from the generators, but also captures waste heat by using the steam to heat local buildings by installing a network of pipes, maximizing the recovered value.

The most commonly used incineration technology is so-called mass burn, where mixed waste is burned in large furnaces on moving, inclined grates. This incineration technology is well established worldwide and proven over many years. The first incineration plants in the UK that burnt mixed

fuel producing steam to generate electricity were built in Nottingham in 1874. Ongoing research and development continues to improve the technology, for example in terms of increasing efficiency, reducing emissions, and lowering capital costs. Fluidized bed technology is thermally more efficient than incineration, but incinerators have not yet been fully developed in the U.K. or U.S. In many municipal waste sites, a plasma arc is used to heat solid waste to very high temperatures (3,000–10,000°C). The product is a hydrogen-rich gas and an inert glassy residue from the nonorganic waste. EfW incineration is not the whole answer, but—as with recycling and composting—when sensibly applied, it can help ensure that the maximum value is recovered from materials that would otherwise be consigned to landfill and lost.

There is a very wide range of municipal or industrial wastes that may be used as fuel. The nature of the waste, and the waste disposal method, will determine the way that energy can be recovered. Dry household, commercial, or industrial wastes can either be burned (combusted) as raw waste, or they may first undergo some sorting or processing to remove some of the waste components that may be recycled separately. Waste incineration is an established way to dispose of wastes. Incineration consumes the biodegradable fraction of the waste, decreases the volume of the waste, and allows for recovery of metals and other potentially recyclable fractions, so that the remaining residue, which must be sent to landfill, is more stable and does not generate potentially harmful emissions such as methane. The heat recovered from waste incineration can be used to generate electricity or used for industrial heat applications. The size of an EfW plant is designed to meet the waste-disposal needs of the community that it serves. Such a plant can typically process between 100,000 and 600,000 tons per year, and from this it can generate up to 40 megawatts of electricity. Power is produced from these wastes by using the steam raised in the combustion process to drive a steam turbine to generate electricity. This technique uses the same proven technologies as those in power plant fueled with fossil fuels such as coal.

Gasification or pyrolysis is one of the later twentieth century technologies that is increasingly being used for waste disposal. It is a thermochemical process in which biomass is heated, in the absence of air, to produce a low-energy gas containing hydrogen, carbon dioxide, and methane, together with tar or ash. The gas can then be used as a fuel in a turbine or combustion engine to generate electricity. Gasification was first developed in the late nineteenth century to produce gas from heated coal or coke. A process developed by Ludwig Mond of the Brunner, Mond & Company (which later became Imperial Chemical Industries, or ICI) was used commercially from 1901 to generate "producer gas" for large industrial furnaces. Gasifiers are now increasingly being developed to accept more mixed fuels, including wastes. The first plants in the U.S. to use gasification to treat municipal solid waste were built in the 1970s, although by 2002 there were no commercial-scale solid waste gasification systems operating in the U.S. Modern gas treatment technology ensures that the resulting gas is suitable to be burned in a gas engine. Gasifiers operate at a smaller scale than a mass-burn EfW plant, and they can also be provided in modular form to suit a range of different scales of operation.

Strict environmental standards now apply in all European countries governing the atmospheric emissions from EfW plants, particularly of heavy metals and dioxins. All energy from waste plants must now meet these standards, which can be achieved through the installation of extensive state-of-the-art gas cleaning systems. An EfW plant managing a guaranteed proportion of residual waste helps to underpin recycling activities by providing a treatment option when markets for recyclables are poor. Moreover, ferrous and nonferrous metals recovery at EfW plants enables many thousands of tons of metal to be recycled. The ash produced from such plants can also be used as an inert aggregate in the construction industry, avoiding the need for disposal of the ash to landfill. Experience in a number of European countries shows that EfW and high recycling rates coexist happily. Indeed, many of the countries achieving high recycling rates also tend to have high-energy recovery rates.

EfW also has an important role in reducing the pollution of carbon dioxide (CO_2), sulfur dioxide (SO_2), and oxides of nitrogen (NO_x) caused by electricity generation from coal, oil, and gas. All power generation processes produce emissions; EfW processes, however, are very well regulated and controlled and, as a result, are far less polluting than fossil sources used for electricity generation. European Directives governing the operation of EfW plants, for example, are far more rigorous and detailed in scope than those applicable to conventional power stations.

More than 60 percent of the heat value of waste is derived from carbon-based renewable biomass sources; that is, wood, paper, and vegetable matter (material that has absorbed CO_2 when it was

growing). When these materials are burned in an EfW facility, the carbon that was trapped in them is rereleased as CO_2. Because there is no net increase in CO_2 however, such a process is CO_2-neutral.

See also **Biomass Power Generation; Electricity Generation and the Environment; Energy and Power**

IAN BURDON

Further Reading

Department of the Environment, Transport and the Regions. *Waste Strategy 2000 for England and Wales*. Her Majesty's Stationary Office, London, 2000.

MacNeill, J.R., Ed. *Something New Under the Sun: An Environmental History of the Twentieth Century*. Penguin, London, 2001.

Petts J. Incineration as a waste management option, in *Waste Incineration and the Environment*, Hester, R.E. and Harrison, R.M., Eds. Royal Society of Chemistry, Cambridge, 1994.

Williams, P.T. *Waste Treatment and Disposal*. Wiley, New York, 1998.

Useful Websites

http://www.wasteguide.org.uk/

National Energy Technology Laboratories, Gasification Technologies: http://www.netl.doe.gov/coalpower/gasification/

UNEP United Nations Environment Program, Division of Technology, Industry, and Economics, Newsletter and Technical Publications: Municipal Solid Waste Management, Regional Overviews and Information Sources: North America: http://www.unep.or.jp/ietc/ESTdir/Pub/MSW/RO/contents_North_A.asp

Power Tools and Hand-Held Tools

While the basic hand tools—hammers, saws, planes, and wrenches—used in construction during the twentieth century changed little from those available for generations, there was a revolution in power tools. Developments in power technology led to the mechanization of tools of all types. Coupled with efforts to use new materials that made tools both lighter and more manageable, construction work became more efficient and cost effective.

Few power tools were available at the beginning of the twentieth century. Those driven by compressed air had their roots in the pneumatic rock drills of the nineteenth century. That technology was easily modified into chipping hammers, paving breakers, and jack hammers. Rivet hammers, developed before 1900, were never more effective than when used to assemble the steel frames of skyscrapers during the 1920s and 1930s.

The evolution of hand-held power tools was dramatic. The first portable electric drill weighed more than 20 kilograms and was in use shortly after 1900. It not only took two men to manage the drill, but a third was needed to control the power supply. A breakthrough in the technology came in 1917 when the Black & Decker Company in the U.S. introduced its first small portable electric drill. Assembled in an aluminum housing, the drill could be held and operated by one worker. Expanding electrification and the miniaturization of electric motors made it possible as well as practical to motorize all sorts of tools. In 1924, the Michel Electric Handsaw Company (later Skilsaw) produced the first portable electric handsaw. This rotary blade device changed the construction industry by dramatically reducing the time needed to cut and fit lumber. Although many of the early developments came from the U.S., Germany was also active in tool design. During the mid-1920s, Gottlieb Stoll produced electric drills and saws. In 1932, the Robert Bosch Company produced an electropneumatic hammer, which was the first in a long line of power hand tools.

The widespread manufacture of small air-cooled internal combustion engines following World War II had a remarkable impact on the construction industry. They made it possible to motorize an even greater variety of tools than ever before. Tools as varied as shovels and wheelbarrows were motorized or self-propelled. One of the first examples of a powered wheelbarrow was introduced in the U.S. by the Bell Aircraft Company in 1948. The operator walked behind and steered the "Bell Prime Mover" as it transported loads of up to 450 kilograms. The tool's usefulness could be broadened by replacing its bucket with either a platform for carrying stacked loads or dozer blade. Similar devices with a wide range of accessories were built in Europe.

Concrete work was improved with the introduction and use of several power tools. A vibrator was devised to remove voids in large pours. The number of workers needed to finish concrete was reduced when the power trowel was marketed in the late 1940s. Traditionally, large expanses of freshly poured concrete were smoothed by workers who pulled wooden or metal floats across its surface. A single laborer using a power trowel could accomplish the same work in less time. Powered by a gasoline engine, the spinning blades of the propeller-like trowel skimmed across and trued the surface of partially set concrete.

The process of setting fasteners in concrete slabs or block walls was simplified with the development

of powder-actuated tools (PATs). Ramset in the U.S. and Hilti in Liechtenstein developed these tools in 1947 and 1948, respectively. Expanding gases from the explosion of a precise powder charge drove a piston against the fastener with such force that it would penetrate concrete.

In the late 1950s and 1960s, nailing was made easier with hand-held nail guns. The gun contained a magazine of nails that were set with the pull of a trigger. These time saving and easy to use devices, powered by compressed air or electricity, were a fast way of joining wooden framing and plywood.

One of the more significant late twentieth century developments was the introduction of cordless battery powered hand tools. This took place in 1962 when the Porter Cable Company in the U.S. marketed its line of "Big 10" cordless drills. Each was powered by 10-volt (hence the name) battery packs worn by its operator. However, the batteries held a charge that lasted no more than 5 or 6 minutes of use. Nonetheless, this new type of tool gave workers previously unknown freedom of movement.

The power for these tools came from nickel–cadmium (NiCad) batteries which were first developed in Germany during World War II. Not only could the cells be readily recharged, they also produced a constant voltage as they discharged. As batteries were improved, recharge time was reduced and the duration of a useful charge increased. Long life and quick recovery made it practical to adapt an increasing variety of tools to battery operation. A significant advancement occurred in the late 1960s when the nickel metal hydride battery was developed. It maintained its charge 50 percent longer than the NiCad batteries it replaced. By the end of the twentieth century rechargeable batteries had improved to the point where they contained energy sufficient to power high-speed circular saws and hand-held planers. The Japanese firm Makita became a leading international manufacturer and innovator of these cordless power tools.

The construction industry benefited from the development of the laser during the 1960s when, in 1968, the rotary laser was used for the first time. That battery-operated tool was used to determine rectilinearity and it soon displaced the optical levels that builders had used for decades. The perfectly straight line indicated by the laser beam had interior and exterior applications. A laser incorporated into the end of a traditional spirit level provided the means for projecting a straight line beyond the reach of the level itself.

During the 1990s the versatility of PAT-type fastener drivers was improved upon with the introduction of a new propulsion system. Unlike PATs that had no more than 10 shots per load, the new gun powered by a cartridge of highly flammable compressed gas had as many as 1200 shots. The interior of the revised fastener gun was similar to an internal combustion engine in which a piston was moved by the explosion of a charge of gas and air ignited by an electric spark.

How tools were used and their impact on the user, led to changes in the design of many handles, grips, and triggers. Concern for the overall weight of tools led to a greater use of plastics and alloys. The distribution of weight within tools led to some overall redesigns in which centers of gravity were repositioned for better balance. The 1990s was a period during which the ergonomics of hand and power tools were scrutinized.

William E. Worthington, Jr.

Further Reading

Allhands, J.L. *Tools of the Earthmover*. Sam Houston College Press, Huntsville, 1951.

Cacha, C.A., Ed. *Ergonomics and Safety in Hand Tool Design*. Lewis, Boca Raton, FL, 1999.

Mcdonnell, L.P. *Portable Power Tools*. Delmar Publishing, Albany, 1962.

Nagyszalanczy, S. *Power Tools*. Taunton Press, Newtown, 2001.

Presentation of Technology

Representations of technology in various media—including expositions, theme parks, fiction, and film—have played central roles in the way human beings have imagined technology and its relationship to progress. Beginning with the scientific revolution in the seventeenth century, when the idea of progress became increasingly secularized and associated with individual and social betterment here on earth, the modern definition of technology as the application of scientific knowledge for human betterment began to emerge. But, as technology became increasingly associated with industrialization in the eighteenth and nineteenth centuries and its negative social consequences, the assumption that better technology would inevitably lead to human progress, came into question. As doubts increased (think of William Blake's "The Songs of Experience") and fed political movements like Luddism, positive representations of technology became increasingly important to the builders of industrializing nation-states who sought to counter growing political opposition to their

authority and power with exhibits that made human progress seem synonymous with industrialization, consumerism, and the globalization of capitalism. Though typically less overtly political in aim and message, presentations of technology in fiction and, later, film frequently echoed and reinforced the optimistic views of the international exhibits. However, novels and movies also offered a means for expressing doubts about the dominant view as they tapped into the persistent current of fear and anxiety about the industrial or postindustrial future that always ran just beneath the surface.

Expositions and World's Fairs

Anxieties about the future became especially acute in mid-nineteenth century England in the aftermath of the 1848 political revolutions in continental Europe that deepened concerns between England's industrialists and their political supporters about the upsurge of Chartist reforms that threatened their prerogatives. To counter political opposition at home and to establish a cultural breakwater against the seas of revolution abroad, England's political authorities, led by Prince Albert, announced plans to organize a "Great Exhibition of the Industry and Works of All Nations." Christened the Crystal Palace Exhibition, the 1851 fair proved tremendously successful in building popular support for the English nation-state and its colonial expansion. The Crystal Palace Exhibition launched the world's fair movement that, over the next century and a half, swept the globe and helped give meaning to technology in the modern and postmodern world. Subsequent exhibitions in France promoted a view of the world that drew sharp distinctions between an industrialized "civilization" and a technologically backward "savagery," thus giving the meaning of progress a definition that was at once technologically and racially inflected. In the U.S., where the devastating industrial depression that began in 1873 inspired waves of political and social upheaval, the 1876 Philadelphia Centennial Exposition helped to build confidence that an industrializing nation could promote progress through economic growth.

However important the equation of "technology equals economic growth equals progress" was to the Victorian era, it was never easy to maintain. In the first half of the twentieth century, two world wars bracketing the Great Depression provided new challenges to sustaining this belief system. Even before World War I ended, political and economic leaders in Europe and the U.S. began laying plans for a new generation of world's fairs that would rebuild Europe as well as the public's faith in scientifically and technologically based progress. Their efforts met with little public enthusiasm until the world's capitalist economies collapsed during the Great Depression. Threatened by the rise of Soviet communism, western capitalists dedicated millions of dollars to revitalizing the world's fair medium as a means toward securing popular faith in capitalism. As the American fairs of the 1930s made clear, representations of technology were central to this effort.

At the 1933–1934 Chicago Century of Progress Exposition, the theme of the fair: "Science Finds, Industry Applies, Mankind Conforms" inspired an artist to sculpt a heroic trinity depicting a robot hunched over and nudging forward figures representing mankind and womankind. At all of the U.S. fairs of the 1930s, representations of labor-saving devices like dishwashers and air conditioners along with new entertainment technologies such as television, held out the promise of a science-based, technologically driven consumerist utopia in America's future if only the existing political and economic order remained essentially intact. The 1939–1940 New York World's Fair brought the era to close with its "Dawn-of-a-New Day" theme and its Futurama and Democracity exhibits that depicted America's arrival at perfection within a lifetime.

The devastation of World War II, especially the Nazi extermination camps and U.S. use of nuclear weapons, made glib assertions about the future very difficult to sustain, but promoters of the 1958 Brussels Universal Exposition pulled out all the stops in their efforts to promote nuclear power as a safe, energy-efficient fuel; as did the promoters of the 1962 Seattle Century 21 Exposition who settled on space exploration as the theme for their fair. Not until Montreal's Expo '67 did a major world's fair suggest that basic assumptions about the meaning of technology and progress needed reconsideration. With its "Man in Control?" subtheme, Expo '67 led to a generation of postmodern international expositions that gave increasing attention to the nuclear and ecological crisis confronting humanity. However, these fairs, dominated by multinational corporations, remain showcases for cultural technologies dedicated less to finding long-term solutions to global problems than to propping up confidence in corporation-run nation-states. Whether the upcoming world's fairs planned for Japan and Shanghai break the mold remains to be seen at the time of writing.

Fiction

Fictional representations of technology became increasingly common in the twentieth century as industrialization and the machine emerged as central and, in the eyes of some, defining characteristics of modern societies. Immediately after their introduction, innovations such as automobiles, radios, and airplanes most frequently figured in novels and stories as creative plot devices or convenient symbols of modernity, though often with little substantial commentary on the broader social or political significance of the technology. Once technologies had lost their novelty, often with astonishing rapidity, they continued to appear without comment in the background of fiction of all types. This ultimate transformation of the new into the unremarkable was perhaps the most telling sign of the centrality of technology to the twentieth century, suggesting that technological change was increasingly viewed as simply an inevitable aspect of life in the modern age.

Among authors whose work commented expressly on the significance of technology in society, views ranged from unalloyed technophilia to dire predictions of machine-driven Armageddon. Many works of fiction elaborated on basic themes that emerged during the previous century as authors first began to ponder the potentially revolutionary effects of industrialization. The British author Mary Shelley's 1818 novel, *Frankenstein*, in which a hubristic scientist creates and loses control of a technological monster was subsequently echoed in countless books and movies that explored the theme of what Langdon Winner has called "autonomous technology": Are humans in control *of* or controlled *by* their machines? In the U.S., the tumultuous process of industrialization inspired both unreasonable hopes and ominous fears. A decade after the Philadelphia Centennial Exposition of 1876, Edward Bellamy's *Looking Backward* (1888) reinforced the idea of industrial utopianism, suggesting that the economic and social dislocations of the Gilded Age could be solved not by turning away from industrialization and technology but by more fully embracing the rationality of the machine. Set in Boston in the year 2000, Bellamy's novel prophesizes that a utopian age of leisure, peace, and unbounded consumer affluence might be achieved if only the productive power of the factory system is properly used for the greater good of society. Though considerably less influential than Bellamy's book (which inspired thousands of Americans to join clubs dedicated to realizing his prophesy), Mark Twain's novel of the subsequent year, *A Connecticut Yankee in King Arthur's Court* (1889), reflected the darker fears of many Americans that Bellamy's work and the industrial expositions sought to allay. In Twain's pessimistic fantasy, a time-traveling Yankee engineer's introduction of modern machinery into Arthurian England brings not leisure and peace but mass destruction and carnage.

Twain's cautionary tale notwithstanding, technological optimism and exuberance remained the dominant fictional theme well into the twentieth century, perhaps suggesting the efforts of industrial and political elites to equate technology, economic growth, and progress were succeeding. In Great Britain, France, and the U.S., the inventor, engineer, and industrialist became popular public heroes, and their quasifictional counterparts frequently played starring roles in popular adventure and romance novels. Tellingly, other writers demoted actual human beings to secondary roles, establishing an enduring genre of fiction (one thread in the complex weave of twentieth-century science fiction) in which science and machines were the real stars. The French novelist Jules Verne pioneered this approach in the previous century with his series of novels envisioning a not-too-distant future of undersea ships, rapid circumglobal air travel, and the human exploration of space. Unapologetically technophilic, Verne's novels gloried in the sheer adventure and excitement of powerful machines while also providing reasonably accurate technical and scientific explanations of their principles. Verne's twentieth century literary progeny found a new home in the pulp fiction science magazines inaugurated in 1926 with Hugo Gernsback's *Amazing Stories*. Graced with vividly colored and imagined illustrations of interplanetary rockets, flying buzzsaws, and mysterious alien cities, the pulp fiction magazines deliberately sought to spark the sense of wonder and awe that historian David Nye calls the "technological sublime." Out of dramatic necessity the stories often explored the multitude of ways in which new technologies could go awry, yet the overall message of the magazines was resolutely optimistic: technology can and will create a better and infinitely more exciting world of tomorrow.

The events of the twentieth century, however, made the pulp fiction promise of progress appear increasingly naïve. The economic misery of worldwide depression, the rise of fascist and totalitarian states, and the appalling mechanical carnage of two world wars forced even the most faithful to question whether technology truly equated with progress. The British author H.G. Wells' novel,

The Shape of Things to Come (1933), envisions a future war of such massive technological destruction that civilization reverts to the conditions of the Middle Ages. Two of the most influential dystopian novels of the century—George Orwell's *1984* (1949) and Aldous Huxley's *Brave New World* (1932)—warned that the exciting new consumer products of the electronic, chemical, and medical industries, might also be used to extend the power of totalitarian states to control every facet of human existence. In Orwell's society of 1984, televisions and surveillance cameras allow "Big Brother" to constantly watch over and propagandize an almost perfectly controlled citizenry, while Huxley's new world government uses cloning and mechanical wombs to rationally manage even the most intimate realms of the human body.

After World War II, atomic weapons and power, digital computers, space rockets, and other new products of the military-academic-industrial complex each inspired a share of adoring fictional treatments. Yet the sheer destructive and transformative power of these new devices also suggested that earlier predictions of future technological disasters might now be coming true. By the time of Walter Miller's 1959 novel, *A Canticle for Liebowitz*, nuclear weapons provided an all-too-plausible means for realizing H.G. Wells' earlier preatomic idea of a massive war that sends civilization reeling back into the Middle Ages. Likewise, Kurt Vonnegut's *Player Piano* (1952) returns to the oft-explored theme of technological unemployment with the difference that the skilled motions of a soon-to-be-unemployed mechanic are now copied into the memory of EPICAC (an obvious reference to ENIAC, the first electronic digital computer). Yet if technology could now realize disaster, perhaps it could also create paradise. Growing public concerns about pollution and overpopulation inspired a new generation of authors to critique technology from a distinctly environmental perspective. Ernest Callenbach's 1975 novel, *Ecotopia*, echoed Edward Bellamy's attempt nearly a century earlier to envision a reasonably realistic alternative to industrial society. Drawing on new ideas from the appropriate technology movement, Callenbach imagines a future nation that deliberately rejects certain advanced technologies as environmentally unsustainable or socially corrosive.

In the final decades of the century, new technologies continued to inspire new fictional treatments. With his 1984 novel *Neuromancer* William Gibson sparked the science fiction subgenre of cyberpunk, dedicated to exploring emerging technologies like virtual reality and global information networks. Perhaps one of the most telling fictional portraits of technology and society in the postmodern age was Don DeLillo's 1985 novel, *White Noise*. Set in a small college town in the middle America of the 1980s, *White Noise* suggests modern technology has brought neither the utopian joys nor the dystopian destruction envisioned by earlier writers. Instead, it has become the largely unnoticed fabric of everyday human existence, a constant "white noise" of babbling TVs, airwaves filled with electromagnetic radiations, and mysterious chemicals that may or may not be slowly killing us. In the postmodern age of technological abundance and unknown risks, DeLillo seems to suggest, age-old human fears and hopes remain unchanged as the latest new thing is embraced and then forgotten, quickly assimilated into the fabric of a society that ultimately remains stubbornly and splendidly merely human.

Film

A technological revolution in their own right, motion pictures proved to be one of the most influential means of exploring the meaning of twentieth-century technology. Early film makers eager to demonstrate the medium's unique ability to portray movement were drawn to the dynamism of trains, autos, and trolleys. Thomas Edison's 1903 western, *The Great Train Robbery*, included the first of what would be a long line of dramatic film fights staged on top of a careening train. The pioneering actor and director Buster Keaton brilliantly developed the comic potential of steamships, automobiles, and trains in *The Navigator* (1924) and *The General* (1926). But other early films offered considerably darker views. The German film *Metropolis* (1926) depicts a future society where the elite few live in luxurious ease while the mass of workers toil ceaselessly below in a sunless subterranean world of dangerous machines. Inspired in part by director Fritz Lang's visit to New York City, *Metropolis* reflects growing concerns about the power of technocratic elites and the fear that mass production was an enslaving rather than liberating force. The American film *Modern Times* (1936) explores similar themes, though director–actor Charlie Chaplin softens his message with a humor completely absent from the darkly serious *Metropolis*. A pointed if ultimately ambiguous critique of Taylorism, Fordism, and the deskilling and subordination of labor, *Modern Times* vividly conveys fears of industrial domination when

Chaplin's character is dragged into a mass of gigantic gears, almost literally becoming a mere "cog in the machine."

In the post-World War II period, the atomic bomb and Cold War tensions with the Soviet Union made unalloyed celebration of technology increasingly difficult to sustain. In Japan the atomic devastation of Hiroshima and Nagasaki provided a subtext to Inoshiro Honda's 1954 film *Godzilla* in which radioactivity from American bomb tests in the Pacific awakes a gigantic reptile that proceeds to crush much of downtown Tokyo. In *Them!*, an American film of the same year, nuclear radiation causes ants to mutate into marauding truck-sized monsters who threaten Los Angeles. Other films of the 1950s dealt more bluntly with the genuine peril of the Cold War nuclear arms race, as in *The Day the Earth Stood Still* (1951) where an alien messenger warns humans not to extend their war-like ways into outer space or risk utter destruction. Abandoning the fantastic all together, *On the Beach* (1959) uses stark realism to offer a sobering tale of the final days of the last survivors of a global nuclear war.

As if the dangers of nuclear Armageddon were not enough, directors explored dozens of less obvious ways in which the countless new postwar technologies might threaten humanity. Stanley Kubrick's homicidal computer HAL in the epic *2001* (1968) suggests the dangers of over-reliance on machines, though the film ultimately portrays advanced technology as the stepping stone to human transcendence. Fears of intelligent machines rebelling against humanity combined with uneasiness about the radical potential of genetic engineering to produce some of the most disturbing representations of technology in the final decades of twentieth century. The cybernetic assassins and guardians of the three hugely popular *Terminator* films (1984, 1991, 2003) and the corporate-created cyber-policeman of *Robocop* (1987) became the latest incarnation of Dr. Frankenstein's monster-machine that rebels against its creator. However, in all these films humanity is ultimately triumphant, either destroying or taming the machine and thus perhaps reaffirming the ultimate rightness of technological progress. An exception was the 1982 film *Blade Runner*, director Ridley Scott's disturbing story of a race of genetically engineered "replicants." Created by a bioengineering corporation as human slaves, the replicants eventually come to be seen as a threat that must be destroyed, even though they prove to be more human—and humane—than their creators in many ways. Given the rapid progress in genetic engineering that resulted in the first map of the human genome in 2000, *Blade Runner* might well be seen as emblematic of a time when the boundaries between the biological and the mechanical or between the human and the machine became increasingly difficult to define.

If films like *Blade Runner* raised troubling questions about technological progress in the last quarter of the twentieth century, technological exuberance and optimism by no means disappeared. The short-lived but influential television drama *Star Trek* (1966–1969) and the subsequent legions of small- and big-screen sequels offered a generally optimistic future in which advanced technology has eliminated human hunger, poverty, and a host of other social ills. Likewise, the immensely popular *Star Wars* saga (begun in 1977) deliberately revived the uncomplicated technological enthusiasm of earlier film and television serials like Flash Gordon. Thus the worldwide popularity of the *Star Trek* and *Star Wars* films may suggest that, wisely or not, many citizens of the late twentieth-century postindustrial and postmodern world ultimately continued to be confident that technology remained an essentially beneficial and beneficent force for human progress.

ROBERT W. RYDELL AND TIM LECAIN

Further Reading

Basalla, G. Keaton and Chaplin: the silent film's response to technology, in *Technology in America: A History of Individuals and Ideas*, Pursell, C.W. Jr., Ed. MIT Press, Boston, 1986.

Findling, J.E. and Pelle, J.D. *Historical Dictionary of World's Fairs and Expositions, 1851–1988*. Greenwood Press, New York, 1990.

Hendershot, C. *Paranoia, the Bomb, and 1950s Science Fiction Films*. Bowling Green State University Popular Press, Bowling Green, Ohio, 1999.

Nye, D.E. *Narratives and Spaces: Technology and the Construction of American Culture*. Columbia University Press, New York, 1997.

Rydell, R.W. *World of Fairs: The Century of Progress Expositions*. University of Chicago Press, Chicago, 1993.

Rydell, R.W., Findling, J.E. and Pelle, K.D. *Fair America: World's Fairs in theUnited States*. Smithsonian Institution Press, Washington D.C., 2000.

Telotte, J.P. *A Distant Technology: Science Fiction Film and the Machine Age*. Wesleyan University Press, London, 1999.

Telotte, J.P. *Science Fiction Film*. Cambridge University Press, New York, 2001.

Tichi, C. *Shifting Gears: Technology, Literature, and Culture in Modernist America*. University of North Carolina Press, Chapel Hill, 1987.

Trachtenberg, A. *The Incorporation of America: Culture and Society in the Gilded Age*. Hill and Wang, New York, 1982.

Printers

Although numerous types of computer printers were developed during the last half of the twentieth century, printing technology emerged during antiquity. The earliest methods of printing included human scribes in Egypt and Greece, but other forms of printing technology existed around the globe including that developed by the Indians and the Mayans. It is China, however, that is notable for its contributions to automatic printing technology. Block printing emerged in China during the ninth century and baked-clay moveable type was introduced in the early eleventh century. In Europe, the Gutenberg press introduced moveable type in the fifteenth century. Moveable type was a revolutionary advancement in printing technology, but it also was the introduction of the telegraph in 1844 that significantly contributed to the development of automatic printing. In fact, the American Wheatstone Automatic Telegraph (1858) and the French Baudot Telegraph (1874) made it possible for telegraph sites to automatically receive messages in Morse code without the need for skilled operators.

In 1910, American engineers Charles and Howard Krum oversaw the installation between New York City and Boston of the first teletypewriter system. Unlike telegraph systems, teletype did not utilize Morse code. Instead the teletypewriter was an input device in which the user could transmit alphanumeric characters typed in one at a time. The system also was equipped with an early teleprinter that was composed of a roll of paper that automatically printed incoming messages. Teletype technology steadily improved, and by 1914, the Associated Press used teletype systems to communicate information to U.S. newspapers. After World War I, these systems were adopted for use throughout the world and remained active for much of the twentieth century. For example, AT&T had 60,000 teletype centers in the U.S. by 1962. The decline of teletype use can be attributed to the introduction of other automated communications systems, including fax machines and email. However, teletype features remain in use today in TTY communication technologies.

Automated printing technology undertook its most significant leap with the advent of the electronic computer. Modern digital computers emerged during the World War II era but reading the output of these early computers was a laborious process. For example, results from the first incarnation of the Manchester Mark I, a computer developed in the U.K., were presented in binary on the face of a display tube—a cathode-ray tube, with bright spots representing 1 and dim spots representing 0. Punched paper tape output was also available. By 1949, however, another British computer, the EDSAC, began displaying its output via an attached teleprinter, to rolls of ordinary paper. Within three years, the American company Remington Rand introduced the first printer designed for use with a computer. These early printers were essentially typewriters attached to a computer; but as computer hardware technology evolved, so did computer printers. In 1971, Gary Starkweather, a researcher at Xerox Palo Alto Research Center, adapted xerography technology in his development of the first laser printer. In the late 1970s, Hewlett Packard began developing inkjet printer technology, which culminated in the introduction of relatively inexpensive color inkjet printers two decades later. During the last 30 years of the twentieth century, the major developments in printing technology emerged primarily from research groups working within the U.S. computer industry.

Printers generally can be categorized as either impact or nonimpact. Like typewriters, impact printers generate output by striking the page with a solid substance. Impact printers include daisy wheel and dot matrix printers. The daisy wheel printer, which was introduced in 1972 by Diablo Systems, operates by spinning the daisy wheel to the correct character whereupon a hammer strikes it, forcing the character through an inked ribbon and onto the paper. Dot matrix printers operate by using a series of small pins to strike a matrix or grid ribbon coated with ink. The strike of the pin forces the ink to transfer to the paper at the point of impact. Unlike daisy wheel printers, dot matrix printers can generate italic and other character types through producing different pin patterns. Nonimpact printers generate images by spraying or fusing ink to paper or other output media. This category includes inkjet printers, laser printers, and thermal printers. Whether they are inkjet or laser, impact or nonimpact, all modern printers incorporate features of dot matrix technology in their design: they operate by generating dots onto paper or other physical media.

Printers are equipped with a print head, which is the device that moves back and forth across the media producing upwards of thousands of dots. There are four types of print heads which can be used: impact, thermal, inkjet, and electrostatic. In addition to the print head, basic printing technology utilizes printer language, bitmaps, and outline fonts to determine the placement of the dots.

Computers use printer language to tell the printer the dot formatting specifications. These commands compress data and manage color, font size, and graphics. Two influential printer languages are page description languages and printer command languages.

Page description language (PDL) is a device-independent language that describes the appearance of a printed page to the printer. Postscript is perhaps the most well-known PDL. Postscript was introduced by Adobe in 1985 for use with laser printers, and it introduced features such as vector graphics. Although commonly used with printers, Postscript can be used with any image-creating device including image setters, screen displays, and slide recorders.

Unlike PDLs, printer command languages (PCLs) are usually proprietary and thus are designed for work with specific models of printers. Hewlett Packard's PCL is a well-known example of such a language. Their PCL originally was for use with dot matrix and inkjet printers, but has since expanded for use with laser printers.

In addition to printer languages, printers can utilize bitmaps and outline fonts to determine dot placement on the paper. Bitmapped fonts are analogous to Gutenberg's moveable type. These fonts are typefaces of a specific size with attributes such as boldface or underline. Most printers come with a limited choice of bitmapped fonts installed into their permanent memory (ROM). However printers also are designed with random access memory (RAM), which permits a user's computer to send a larger selection of bitmapped fonts to the printer's font library. Printer language enables the computer to send the printer the directions for what bitmap table to use, and the printer uses that bitmap to generate the desired typeface. A limitation of bitmapped fonts stems from the fact that they are not scaleable since they are digital representations of typeface. Since bitmaps are composed of a set size or a limited set of sizes, these printers do not have the ability to completely control the resultant generated image. Printers that use outline fonts are much more functional. Outline fonts are scaleable and are used with Postscript and PCL printer languages. The printer language directs the printer to treat all output as if it were a graphic instead of a typeface. This functionality generates output that is more attractive and versatile than that produced by bitmapped fonts.

Early computer printers were developed for use by government agencies, large corporations, and universities. However, since Centronics Data Corporation introduced the first dot matrix printer in 1970, printers have become increasingly smaller and affordable. In addition, their operation has become increasingly user-friendly. Much of the crossover from institutional to individual use can be illustrated by personal printing developments in the 1980s. In 1984, Hewlett Packard began marketing both personal inkjet and laser printers for use in homes and smaller offices. Inkjet printers are equipped with an ink-filled print cartridge that is attached to the print head. The cartridge contains nozzles through which droplets of ink are sprayed onto paper to generate an image. Hewlett Packard had been working with inkjet technology since the late 1970s, and their innovations included miniaturizing inkjet printers and reducing the level of printer noise. The quality of inkjet output was superior to dot matrix which explains why dot matrix began to lose its appeal to consumers. A variation of the inkjet printer is bubble jet technology. Where inkjet printers use non-heating crystals to generate its output, bubble jet printers use heating elements to shoot ink from nozzles onto the paper.

The 1991 introduction of color inkjet printers by Hewlett Packard is a significant contributing factor as to why inkjet technology retained its popularity at the close of the twentieth century. Color inkjet printers are equipped with four ink cartridges—magenta, cyan, yellow, and black—attached to its print head. These printers use thermal technology to generate images on a wide variety of physical media. Each ink cartridge possesses a thin resistor through which an electrical current flows; this heats a thin layer of ink to more than 480°C for several millionths of a second. The ink is then rapidly sprayed onto the printing medium.

Laser printers are high-resolution printers that use a version of electrostatic reproduction or xerography technology found in copying machines to fuse images onto paper. Laser printers essentially operate by reflecting a laser beam from a spinning mirror to attract ink or toner from a rolling drum onto targeted areas of the paper. As laser printers generate over 300 dots per inch, consumers were able to afford typeset-quality output. Laser printers, and personal computer printers in general, illustrate one of the more important trends in printing technology: the role played by users. Printing technology during the last half of the twentieth century offered users increased opportunities to undertake their own print jobs. Before the development of personal computer printers, clerical staff and other individuals were forced to send their printing jobs elsewhere. Computer printers now offer the

means to quickly generate high-quality, independently produced newsletters, business cards, brochures, and so on. With the advent of digital cameras at the end of the twentieth century, consumers could begin developing their own digital photographs. Although automatic printing was originally directed for use by governments and corporations, consumers were able to adopt this technology. Computer printers have revolutionized printing in much the same way as that of the moveable type.

See also **Computers, Personal; Electronic Communications; Fax Machine**

JULIA CHENOT GOODFOX

Further Reading

Campbell-Kelly, M. and Aspray, W. *Computer: A History of the Information Machine*. Basic Books, New York, 1996.

Hiltzik, M. Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age. Harper Business, New York, 2000.

Lang, H. *A Phone of Our Own: The Deaf Insurrection Against Ma Bell*. Gallaudet University Press, Washington D.C., 2002.

Lubar, S. *InfoCulture: The Smithsonian Book of Information Age Inventions*. Houghton Mifflin, Boston, 1993.

Packard, D. *The HP Way: How Bill Hewlett and I Built Our Company*. Harper Business, New York, 1995.

Standage, T. *The Victorian Internet: The Remarkable Story of the Telegraph of the Nineteenth Century's On-Line Pioneers*. Berkley Publishing Group, New York, 1999.

White, R. *How Computers Work*, 6th edn. Que, Indianapolis, 2002.

Useful Websites

The Associated Press History: http://www.ap.org/pages/history/history.html

American Telephone &Telegraph History: http://www.att.com/history

Processors for Computers

A processor is the part of the computer system that manipulates the data. The first computer processors of the late 1940s and early 1950s performed three main functions and had three main components. They worked in a cycle to gather, decode, and execute instructions. They were made up of the arithmetic and logic unit, the control unit, and some extra storage components or registers. Today, most processors contain these components and perform these same functions, but since the 1960s they have developed different forms, capabilities, and organization. As with computers in general, increasing speed and decreasing size has marked their development.

The early computing machines of the twentieth century relied on various and mechanically complex ways to handle data. The first digital computers, such as the ENIAC built for the U.S. Army and completed in 1945, relied on processing units constructed from thousands of thermionic valves (or vacuum tubes) plugged into connectors (see Computers, Early Digital). Valves were connected in circuits and transferred electronic signals to enable mathematical and logical operations. The thermionic valve increased the speed of a computer's calculations. However, the valves required large amounts of power and were very expensive; they were approximately the size of small light bulbs and made computers very large and difficult to maintain and operate.

The data processing of the ENIAC was hampered not only by the fragility, size, and cost of the valves but also by its inability to store a program and data. In 1945, mathematician John von Neumann synthesized the research he and his colleagues conducted on the ENIAC and outlined the construction for a new computer in his seminal paper *A Draft Report on the EDVAC*, in which he proposed the stored program concept as an answer to the ENIAC's computing problems. The paper also described the basic components and functioning of the processor. While this paper articulated the operating concepts of the modern processor, its physical manifestation would take several more years of work and the use of valves would dominate computer construction into the late 1950s.

The EDVAC processor required a circuitry based on binary logic. Precedents for computing machines based on binary logic existed earlier but were not widely known. Machines built in the late 1930s and early 1940s by pioneers such as Konrad Zuse, Alan Turing, John Atanasoff and Clifford Berry employed a binary system. Claude Shannon, a researcher at the Massachusetts Institute of Technology, also noted the applicability of a computer's system of ones and zeros to the Boolean logic values of TRUE and FALSE. His 1938 paper, *A Symbolic Analysis of Relay and Switching Currents*, was an analysis of this relationship between computer logic and Boolean algebra, a mathematical system devised in the mid-nineteenth century by British mathematician, George Boole.

Binary logic offered a more efficient construction for computer processors. In the first computers, "logic gates," which were groups of valves in

early computers and transistors in later ones, were placed to form circuits and operate together under a binary system in which information was transmitted in ones and zeros. The information traveled via electric current and the voltage of the current determined whether the signal was a one or a zero. The gates received the signals and according to their configuration would output a new signal. The complex circuitry formed by these gates allowed computers to process instructions given by the user. Most early processors contained sets of circuits known as accumulators, registers and control units. Accumulators performed simple arithmetic and stored sums; registers provided temporary storage for data and instructions; and control units directed the processor's operations.

Basing their work on these early processing units, many designers in the 1950s and 1960s developed processors for the large computer companies such as IBM, Honeywell, General Electric, Burroughs, Univac, and Digital Equipment Corporation. However, the physical form and capabilities of the computer processor were most dramatically changed with the invention of the transistor and the integrated circuit. In 1948, at Bell Laboratories in New Jersey, the work of three physicists, John Bardeen, Walter H. Brattain, and William Shockley resulted in the production of the first transistor. It was, however, nearly a decade before transistorized computers would appear for commercial use. The U.S. military supported the construction of the first fully transistorized computer in 1952. Built by Bell Laboratories, the computer was named TRIDAC (transistorized digital computer). Due to the early transistor's cost and unreliability, it was not until the late 1950s that companies devised more sturdy devices and built commercial computer processors using transistors.

Following the construction of commercial transistorized computers, research continued on making electronic components smaller and faster. In 1959, Jack Kilby at Texas Instruments and Robert Noyce at Fairchild Semiconductor, working separately to improve microelectronics and circuit design, invented and filed patents on a device called the integrated circuit or "chip." Integrated circuits were made up of transistors etched onto the surface of a semiconductor, most commonly a thin wafer of silicon. Processors were now made up of groups of integrated circuits placed onto printed circuit boards. Integrated circuits improved the speed and decreased the size of computer processors. After the introduction of an improved semiconductor called MOS (metal oxide semiconductor), several engineers and researchers worked to fit processor components onto a single integrated circuit. This process resulted in the single-chip central processing unit (CPU). The single-chip CPU heralded a new generation of computer processors and the development of the personal computer.

Intel, a company founded by former integrated circuit makers Robert Noyce and Gordon Moore, commercially produced the first single-chip CPU or microprocessor in 1971. This microprocessor, named the 4004 was designed and constructed by Intel employees Ted Hoff, Stan Mazor, and Federico Faggin. The 4004 included 2300 transistors and contained the registers and arithmetic and control units of a basic general-purpose computer processor. It performed 600,000 instructions per second.

Increases in speed were constant after the introduction of the Intel 4004 and several companies competed with Intel to produce smaller and faster microprocessors. In accordance with the popularly termed "Moore's law," in which Gordon Moore observed that approximately every eighteen months after the invention of the integrated circuit the number of transistors on a circuit doubled, the number of transistors in a microprocessor increased significantly with each new model. For example, the Pentium 4, introduced by Intel in 2000, contained 42million transistors and performed 2 billion instructions per second. These increases in processor speed, in conjunction with improvements in computer memory, allowed more and larger software applications to run on a computer. The decreasing costs of manufacturing processors and their increasing compactness has made possible very powerful personal computers.

See also **Computer Memory; Computers, Early Digital; Computers, Mainframe; Computers, Personal; Computers, Uses and Consequences; Integrated Circuits; Transistors; Valves/Vacuum Tubes**

JESSICA R. SCHAAP

Further Reading

Aspray, W. *John von Neumann and the Origins of Modern Computing*. MIT Press, Cambridge, MA, 1990.
Ceruzzi, P. *History of Modern Computing*. MIT Press, Cambridge, MA, 1998.
Kilby, J.S. Invention of the integrated circuit. *IEEE Trans. Electron. Devices*, 23, 648–654, 1976.
Noyce, R. and Hoff, M. A history of microprocessor design at Intel. *IEEE Micro*, 1, 8–22, 1981.

Rojas, R. and Hashagen, U., Eds.*The First Computers: History and Architectures*. MIT Press, Cambridge, MA, 2000.
Smith, R.E. A historical overview of computer architecture. *Ann. Hist. Comp.*, 10, 4, 277–303, 1989.

Prospecting, Minerals

Twentieth century mineral prospecting draws upon the accumulated knowledge of previous exploration and mining activities, advancing technology, expanding knowledge of geologic processes and deposit models, and mining and processing capabilities to determine where and how to look for minerals of interest. Geologic models have been developed for a wide variety of deposit types; the prospector compares geologic characteristics of potential exploration areas with those of deposit models to determine which areas have similar characteristics and are suitable prospecting locations. Mineral prospecting programs are often team efforts, integrating general and site-specific knowledge of geochemistry, geology, geophysics, and remote sensing to "discover" hidden mineral deposits and "measure" their economic potential with increasing accuracy and reduced environmental disturbance. Once a likely target zone has been identified, multiple exploration tools are used in a coordinated program to characterize the deposit and its economic potential.

The field of geophysics uses physical property measurements of the earth's surface and subsurface to locate mineral deposits. Taken as a whole, electrical and electromagnetic methods represent a large class of geophysical methods, measuring current flow, electrical potential (voltages), and electromagnetic fields. Electrical potential measurement can be traced back to the mid-1800s when Robert Fox of Cornwall, England, first used the self-potential (SP) electrical conductivity method to detect extensions of known copper deposits in the 1830s. By the beginning of the twentieth century, electrical conductivity in rock and soils was widely recognized, sonic wave transmission was understood, and magnetic susceptibility and radioactivity were recognized. Conrad Schlumberger (France) discovered the principle of induced polarization (IP) in 1912 by noticing that soil resistivity measurements varied with frequency for some subsurface materials (e.g., clays and pyrite), but not others. By the 1940s, Canadian and later American and European geophysicists were experimenting with active (systems with both source and receiver) and passive (receiver only) electromagnetic (EM) surveying techniques, using both ground-based and aircraft-based platforms, to differentiate highly conductive massive sulfide orebodies from their less conductive host rocks. EM surveying began to flourish after World War II with greater availability of pilots and aircraft, increasing global demand for minerals, and the rise of the integrated mining and exploration company, which could afford to fund such expensive methods. The U.S. Geological Survey and the U.S. Navy conducted the world's first full-scale airborne magnetic survey in 1945.

Over the next 25 years, EM surveys proved useful in the discovery of massive volcanogenic sulfide base-metal deposits and diamondiferous kimberlite pipes, to the point that the frequency of discovery of such deposits dramatically increased. Canadian prospectors made extensive use of EM techniques in their exploration of the Canadian Shield, with the result that Canada became a world-class producer of base metals by the end of the century. Aircraft-borne geophysical systems had other long-term impacts; observation of magnetic banding in the Atlantic seafloor was a key element leading to the development of the concept of plate tectonics during the late 1960s and the early 1970s.

Many techniques routinely used in mineral exploration were originally developed for use in hydrocarbon or mineral-fuels exploration. During the 1930s, geophysicists (nicknamed "doodlebuggers" for the devices used to locate underground water, gas, or ores) developed the first primitive dynamite-source and inductive-coil-response seismic methods in their search for oil and gas. By the 1940s, airborne radiometric surveys had been conducted by the uranium industry to remotely map "radioactive" rocks. Initial experiments with measuring the strength of the gravity field of hydrocarbon-rich salt domes required development of mechanical–optical systems that could detect changes in the earth's gravitational field as small as one part in 10 billion. By the 1950s, methods to measure gravitational field differences reflecting the density contrast between mineralized areas and surrounding material had been expanded to many mineral deposit types. Chromite, iron, and massive sulfide deposits produce positive density contrasts. Mineral deposits that produce negative contrasts include gypsum, potash, and salt. By the end of the century, telemetry and laser ranging techniques were sufficiently advanced that airborne gravimetric systems were being used in regional land and ocean surveys. Although seismic three-dimensional signal processing and modeling is extensively used in oil and gas exploration, applica-

tion to nonfuel mineral exploration is currently limited.

Remote sensing, broadly defined as obtaining information about an object without coming into physical contact with it, has been around since Galileo first used a telescope to observe celestial objects in 1609. The photographic camera has been a prime remote sensor for more than 150 years, since its development in France by Jacques Daguerre and Nicephore Nepce in 1839. Aerial photography, used initially from tethered balloons during World War I, began to play a part in mineral exploration in the 1920s, but did not come into prominence until after World War II. Improvements in helicopters and development of spacecraft-mounted cameras during the latter half of the twentieth century permitted exploration programs to operate in remote regions. With the advent of earth satellites with reflectance imaging systems (such as the Landsat satellite series first launched by the U.S. in 1972), remote sensing began to use instrumentation that indirectly detected minerals by their contrasting brightness levels (air photos) and spectral reflectance measurements. By the 1990s, hyperspectral scanners such as the airborne visible/infrared imaging spectrometer (AVIRIS) operated by the National Aeronautics and Space Administration (NASA) could record up to 200-plus spectral channels simultaneously, improving the instrument's ability to identify individual minerals or rocks (Figure 16). Remote sensing analysis is currently used to characterize geomorphology, lithology, mineralogy, and structure. When processed results are compared with mineral deposit models, prospecting targets can be inferred and field checked. Confidence in target selection can be increased by integrating remotely sensed results with geophysical data, geologic maps, geochemical sampling, topography, and other data.

In addition to the acquisition of various physical properties data, interpretation of geophysical and remotely sensed data made great strides during the second half of the twentieth century. Increases in computational power and data storage capability led to new ways of modeling data and the development of new analysis methods to provide information not apparent in initial imaging of raw data. Powerful imaging algorithms and fast, high-resolution displays allow prospectors to generate

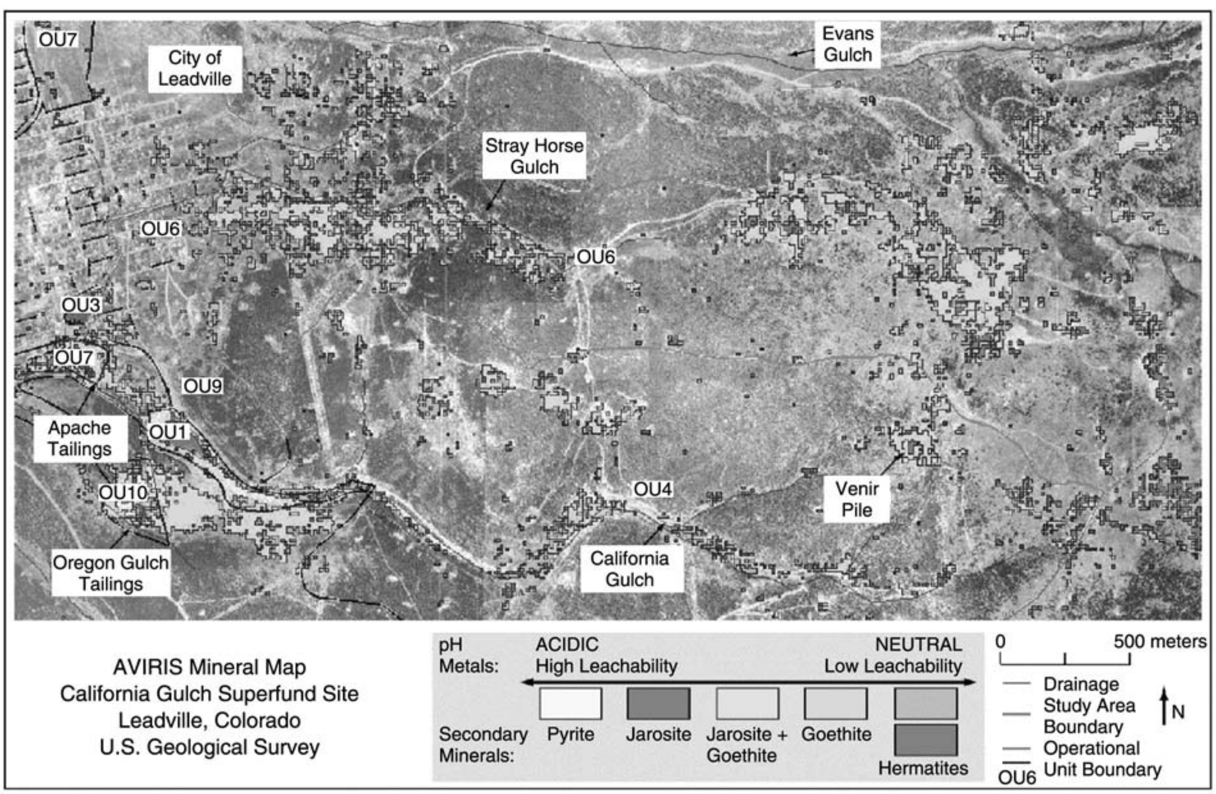


Figure 16. Airborne visible/infrared imaging spectrometer (AVIRIS) mineral map from a hyperspectral scanner operated by the National Aeronautics and Space Administration (NASA).

three-dimensional images of subsurface rocks and structures. Computers allow field teams to transmit data instantaneously, and speed up acquisition and analysis of large volumes of data.

Exploration geochemistry applies chemical principles, properties, and measurements to the discovery, delineation, and development of mineral resources. The evolution and increased contribution of exploration geochemistry to mineral discovery during the twentieth century has been primarily dependent on the increase in demand for and diversity of materials used by society, the decrease in the number of mineral deposits exposed at the surface, and adaptation of technological advances to mineral exploration and extraction.

Trace element analysis as a supplement to visual observation was first used in the mid-1930s when Victor Goldschmidt and J.H.L. Vogt (Scandinavia) and Vladimir I. Vernadsky and Alexander E. Fersman (USSR) independently carried out experiments using spectrographic analysis of soils and plants as a prospecting method. The development of dithizone (a colorimetric chemical reagent) for rapid chemical analyses in the mid to late 1940s has been attributed as one of the most important factors for the subsequent development of geochemical prospecting methods in North America. H.E. Hawkes, T.S. Lovering, and Bloom (U.S.) initially developed colorimetric field analytical techniques in 1947; during the 1950s, J.S. Webb (Great Britain) and other international researchers enhanced these techniques. Alan Walsh (Australia) introduced atomic absorption spectrometry in 1955 and subsequent research by Ward (U.S.) led to the rapid, accurate, sensitive and relatively interference-free analysis of many materials of interest in minerals exploration.

Another key advancement in exploration geochemistry was the development of inductively coupled plasma-mass spectrometry (ICP-MS) during the mid-1980s. This technique allows for cost-effective detection of many solid materials at parts per billion (ppb) or water at parts per trillion levels. The method formed the impetus not only for renewed interest in hydrogeochemistry in exploration, but also the renaissance in selective and partial extraction techniques used to assess processes that control metal migration in fluids. Since less than 1 percent of the metal may be retrieved by such extractions, accurate determination of ppb metal concentrations by ICP-MS technology is essential. As mineral exploration increasingly focuses on subsurface resources, measurement of gases (e.g., carbon dioxide, helium, radon, sulfur, and volatile mercury) in soils and the atmosphere became important. The dramatic increase in diamond exploration during the last decade was aided by the geochemical identification of specific indicator minerals (such as the G10 garnet) that are directly associated with diamondiferous deposits.

Although geochemical and geophysical prospecting can suggest the presence of anomalous mineral occurrences, delineation of the geometry, quality, and size of mineral prospects may best be accomplished by exploration drilling. Major strides in exploration drilling techniques came about with the need to delineate deeper and more geologically complex mineral deposits. The diamond drill, developed by Georges Auguste Leschot (France) in the 1860s, is considered one of the most important drilling tools. This early device used steel bits set with natural diamonds, capable of penetrating hard rock. Drilling was advanced with improved, more efficient drill design, and engine technology during the early part of the twentieth century. With the development of tungsten carbide bits in the 1950s, diamond core drilling advanced. By the 1970s, drill bits using more advanced alloys and impregnated with manufactured diamonds became the industry standard.

See also **Satellites; Environmental Monitoring**

DAVID R. WILBURN, DONALD I. BLEIWAS, KAREN D. KELLEY, K. ERIC LIVO, AND JEFFREY C. WYNN

Further Reading

Bloom, L.B. and Hall, G.E.M. Advances in exploration geochemistry; successful application of selective extraction techniques. *Explor. Min. Geol.*, 6, 3, 281, 1997.

Boyle, R.W., Ed. *Geochemical Exploration*. Canadian Institute of Mining and Metallurgy (CIM) Special Volume 11, Montreal, 1971.

Cox, D.P. and Singer, D.A., Eds. *Mineral Deposit Models, U. S. Geological Survey Bulletin 1693*. U.S. Government Printing Office, Washington D.C., 1992.

Govett, G.J.S. Geochemical remote sensing of the subsurface, in *Handbook of Exploration Geochemistry*, vol. 7. Elsevier, Amsterdam, 2001.

Hanna, W.F., Ed. Some historical notes on early magnetic surveying in the U.S. Geological Survey, in *Geologic Applications of Modern Aeromagnetic Surveys, U.S. Geological Survey Bulletin 1924*. U.S. Government Printing Office, Washington D.C., 1987.

Knepper, D.H., Jr. Mapping hydrothermal alteration with landsat thematic mapper data, in *Remote Sensing in Exploration Geology, 28th International Geological Congress, Field Trip Guidebook T182*, Keenan Lee, Ed. American Geophysical Union, Washington D.C., 1989.

Levinson, A.A. *Introduction to Exploration Geochemistry*. Applied Publishing, Wilmette, IL, 1974.

Lowe, C., Thomas, M.D., and Morris, W.A., Eds. *Geophysics in Mineral Exploration: Fundamentals and Recent Developments*, short course notes vol. 14.

Geological Association of Canada, St. John's, Newfoundland, 2000.
Rencz, A.N. and Ryerson, R.A., Eds. *Manual of Remote Sensing, Volume 3: Remote Sensing for the Earth Sciences*. Wiley, London, 1999.
Rose, A.W., Hawkes, H.E., and Webb, J.S. *Geochemistry in Mineral Exploration*. Academic Press, New York, 1979.
Rowan, L., Abrams, M., Clark, R., Goetz, A., Kruse, F., Lee, K., Lyon, R., Peters, D., Podwysocki, M., and Raines, G. Mineral resources, in *Airborne Remote Sensing for Geology and the Environment—Present and Future: U. S. Geological Survey Bulletin 1926*, Watson, K. and Knepper, D., Eds. U.S. Government Printing Office, Washington D.C., 1994.
Sharma, P.V. *Geophysical Methods in Geology*. Elsevier, Amsterdam, 1986.
Swayze, G.A., Smith, K.S., Clark, R.N., Sutley, S.J., Pearson, R.M., Vance, J.S., Hageman, P.L., Briggs, P.H., Meier, A.L., Singleton, M.J. and Roth, S. Using imaging spectroscopy to map acid mine waste. *Environ. Sci. Technol*., 34, 47–54, 2000.
Telford, W.M., Geldart, L.P. and Sheriff, R.E. *Applied Geophysics*. Cambridge University Press, Cambridge, 1990.
Ward, S.H., Ed. *Geotechnical and Environmental Geophysics*. Society of Exploration Geophysicists, Tulsa, Oklahoma, 1990.

Useful Websites

Boyd, T.M. *Introduction to Geophysical Exploration*: http://www.mines.edu/fs_home/tboyd/GP311/introgp.shtml
Pees, S.T., *Oil History*: http://www.oilhistory.com/pages/diamond/inventor.html
Short, N.M., Sr. *The Remote Sensing Tutorial*, National Aeronautics and Space Administration: http://rst.gsfc.nasa.gov
Wilburn, D.R., Goonan, T.G. and Bleiwas, D.I. Technological Advancement: A Factor in Increasing Resource Use, U.S. Geological Survey Open File Report 01-197: http://pubs.usgs.gov/openfile/of01-197

Psychiatry, Diagnosis and Non-Drug Treatment

Psychiatry is the medical specialty for diagnosing and treating mental, emotional, and behavioral disorders. During the twentieth century, psychiatry adapted and abandoned several technologies in the treatment of mental illness with the notable exception of electroconvulsive therapy, which from its invention in 1938 has remained an important treatment for mental illness.

At the beginning of the twentieth century, there were few nondrug treatments available in psychiatry. Psychotherapy was available for the well-to-do, but others went untreated or were consigned to asylums that were little better than prisons.

One method used for the first two decades of the century was electrotherapy. Electrotherapeutic devices used static electricity, induction coils, or direct current electricity drawn from simple batteries ranging in size from small wooden boxes to the room-sized static electric machines. By passing an electrode, or wand, along the spine and nape of the patient's neck, physicians claimed to cure intractable neuroses such as neurasthenia or hysteria. In other cases patients were seated in front of giant static electrical machines and received a static wind. Electrotherapy fell out of favor after World War I when psychiatrists redefined mental illness and the symptoms clustered under neurasthenia and hysteria were reclassified as psychodynamic rather than physical ailments. Throughout the twentieth century variations of electrotherapy survived as part of physical therapy.

Hyperthermal cabinets, used from the late 1930s through to the early 1950s, were also once part of the psychiatric treatments. The idea for fever therapy, as it was known, came from Julius Wagner von Jauregg who injected patients with blood-borne malaria to induce high fevers and successfully treated general paresis. Rather than risk the side effects of malaria, hyperthermal cabinets were manufactured and sold to hospitals as a safer, more controlled means of raising the body's temperature. The devices, shaped much like the iron lung, encased the patient below the neck and raised the body to temperatures above 106°F (41.1°C). Despite clinical success on both schizophrenia and depression, this treatment was abandoned with the advent of electroconvulsive therapy.

Electroconvulsive therapy, or ECT, was invented in 1938 by Ugo Cerletti who sought to replicate the convulsive therapy pioneered by Ladislas von Meduna. Meduna believed that epilepsy was antagonistic to mental disease and so contrived a convulsive medication called metrazol or cardazol which sent patients into convulsive fits that seemed to cure schizophrenia. Cerletti's research team set about to find a technologically cleaner and simpler method to create convulsions and succeeded in 1937 when Cerletti's assistant Lucio Bini invented a simple electroconvulsive machine. The machine ran on a standard household alternating current of 125 volts and sent electricity through the brain using a pair of calipers attached to the temples. In April 1938 Cerletti gave ECT to a mental patient and his published findings confirmed his success.

In the U.S., machines were first built by psychiatrists themselves and had several common features. Machines tended to be small devices encased in wood with a glass-covered meter that

gave the voltage of the electricity, while another dial allowed the physician to increase the voltage, and a button that when pushed began or ended a treatment. "Home made" devices of wood gave way to metal and a number of manufacturers competed for the practitioners by including a variety of designs and features between 1940–1960. Lektra, Medcraft, Offner, and Reiter were important manufacturers of ECT machines throughout this period and each added technical refinements. The size could vary from 13 by 8 centimeters to that of a large suitcase. The standard models incorporated instruments that could read the patient's electrical resistance and provide a slow rise in current (or glissando); some included so-called reverse glissando because they started with a higher current that stepped down gradually. There were also machines that used the so-called brief stimulus method that delivered electricity in short bursts or waves which presumably lessened the prevalent side effects such as marked memory loss. It was not uncommon in the 1950s to find machines that would also provide what was called electrosleep or electronarcosis. This was a subconvulsive dose of continuous electricity that led to unconsciousness without convulsion. Some physicians used this a treatment for mental or nervous disorders, while others used it to put patients to sleep prior to administering ECT.

ECT fell from favor after the advent of psychopharmacology, and beginning in the 1960s, manufacturers abandoned the field one by one. In the 1980s Medcraft, now called Hittman-Medcraft, was the only original manufacturer remaining in the U.S. However, the psychiatrist Paul Blachy modified a machine in 1973 to incorporate an electroencephalogram (EEG) and electrocardiogram (EKG) to monitor patient's vital signs before and after the treatment. He called his design the monitored electroconcovulsive therapy apparatus (MECTA) and began manufacturing his machines under the MECTA name in the mid-1970s. In the mid-1980s, Richard Abrams and Conrad Swartz combined their efforts to create Somatics, which is the leading manufacturer and seller of ECT devices (particularly its Thymatron line). The new devices incorporated computer chips to provide a number of read-outs and deliver precise amounts of electricity at precise time intervals. Although using higher voltage, some estimates suggested 180 or more volts, these devices cycled at faster rates and had shorter pulses of electricity over a longer duration. Devices constructed in the 1980s also allowed physicians to set the machine according to the age of the patient, an important factor in determining the body's electrical resistance. With the advent of so many new features, the devices have become larger and approximately the size of a large stereo receiver. ECT, which in the 1960s seemed destined for the same end as electrotherapy, hydrotherapy, and fever therapy, made a dramatic resurgence in the 1980s and is particularly useful for medically resistant forms of depression. Estimates suggest as many as a 100,000 treatments are given annually.

The popularity of ECT technology has inspired the use of the experimental treatment known as transcranial magnetic stimulation (TMS), which employs some of the ideas of electrotherapy. In TMS an electromagentic coil is placed on the scalp and an electric current of high intensity is turned on and off through the coil. This creates a magnetic field lasting from 100–200 microseconds, which is repeated in a process known as rTMS. Clinical reports show efficacy among treatment-resistant depressions. Unlike ECT, which employs a general electrical charge through the head, rTMS can be situated to deliver electricity to a more specific region of the brain. Side effects such as headache and memory loss are lessened, patients remain awake throughout the treatment and, though still not approved by the Food and Drug Administration (FDA) as a treatment for depression, this technology remains promising.

The last major medical breakthrough in twentieth-century psychiatry may well be the introduction of vagus nerve stimulation (VNS) to treat depression and anxiety. In the late 1990s psychiatrists began employing this technique, which involves implanting a pacemaker-sized device into the chest wall. The device delivers a short pulse of electricity to the vagus nerve located in the neck, which in turn stimulates a series of nerves that seem to be involved in brain chemistry. Since 1998 clinical results have demonstrated the promise of this new technique.

See also **Electroencephalogram (EEG); Psychiatry, Pharmaceutical Treatment.**

TIMOTHY W. KNEELAND

Further Reading

American Psychiatric Association. *The Practice of ECT: Recommendations for Treatment, Training, and Privileging*. American Psychiatric Press, Washington D.C. 1990.

Braslow, J. *Mental Ills and Bodily Cures: Psychiatric Treatment in the First Half of the Twentieth Century*. University of California Press.

George, M., Harold Sackein, *et al.* vagus nerve stimulation: a potential therapy for resistant depression. *Psychiatr. Clin. N. Am.*, 23, 4, 757–782, 2000.

George, M., Lisanby, S.H. and Sackeim, H. Transcranial magnetic stimulation: applications in neuropsychiatry. *Arch. Gen. Psychiatr.*, 56, 300–311, 1999.

Kneeland, T. and Warren, C. *Pushbutton Psychiatry: A History of Electroshock in America*. Praeger, West Port, CT, 2002.

Valenstein, E. *Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Treatments for Mental Illness*. Basic Books, New York, 1987.

Psychiatry, Pharmaceutical Treatment

Psychiatry, the medical specialty that deals with emotional and behavioral disorders, underwent a profound transformation in the mid-twentieth century with the development of psychopharmacology, the use of drug therapy to treat specific mental illnesses ranging from schizophrenia to depression and anxiety. Following the serendipitous discovery of drugs such as chlorpromazine in 1952, researchers examined the specific effects of these drugs, in turn spawning a revolution in the scientific understanding of the brain and the role of neurotransmitters in the expression of mood and behavior.

Prior to the twentieth century, private practice and asylum physicians used a combination of drugs such as opiates, barbiturates, bromides, chloral hydrate and hyoscine to treat the mentally ill. These nonspecific drugs were used to sedate agitated patients or elevate the mood of the melancholic. Although these attacked the symptoms of the illness, they were adjunct to other treatments that were seen as curative. During the first half of the twentieth century, psychiatric treatments incorporated drug regimens to alleviate the symptoms of ailments classified as either psychosis or neurosis. One such regimen was sleep therapy, or prolonged sleep, popularized by Jacob Klaesi in 1920. Using various bromides or barbiturates, Klaesi put patients to sleep for days at a time and found that prolonged sleep decreased the symptoms of the illness but did not cure patients. In their search for a cure some psychiatrists turned to shock or convulsive therapies. Based on the notion that epileptic patients were immune to schizophrenia, psychiatrists tried a host of convulsive agents to cure that schizophrenia. Beginning in the 1930s Ladislas Meduna pioneered the use of the chemical agent pentylenetetrazol, also known as Metrazol or Cardiazol, to cause patients to convulse. Similarly, Manfred Sakel developed insulin therapy, another shock or convulsive therapy. In this treatment, popular from the 1930s to the mid-1950s, synthetically produced insulin was employed over the course of hours to induce comas and convulsions that seemed to alleviate the symptoms of schizophrenia. The development of electroconvulsive therapy (ECT) and psychopharmacology made Metrazol and insulin therapy obsolete after 1950.

Drug therapies appeared as a byproduct of research in the chemical and pharmaceutical industry. Chlorpromazine (CPZ), better known by the trade name Thorazine, had been explored for its properties as a dye in the late nineteenth century and for its antihistamine properties in the first half of the twentieth century. Its sedative properties were noted and several clinicians including French surgeon Henry Laborit, as well as Jean Delay and Pierre Denker, were to independently confirm its ability to calm agitated mental patients and alleviate the symptoms of schizophrenia. Thus was born a new class of drugs known as neuroleptics in Europe and antipsychotics in North America. Chlorpromazine set the pattern for the psychopharmacological revolution and followed the life-cycle taken by almost every other pharmaceutical innovation: drugs were introduced with high expectations and enthusiasm, but after the emergence of side effects they were derided and in some cases abandoned.

After its introduction in North America by SmithKline & French in 1953 and its approval for use in the U.S. by the Food and Drug Administration (FDA) in 1954, CPZ became a standard therapy in the state mental institutions leading to the dramatic decline of inmates from 500,000 to 150,000 within a short period of time. However, within a decade severe side effects that created a Parkinson-like disorder known as tardive dsykinesia were discovered and alternative drugs were sought.

The initial success of CPZ led to further research on histamine blockers and in 1957, Roland Kuhn would slightly alter the chemistry of CPZ and create imipramine hydrochloride, (Tofranil), a drug found to be useful in the treatment of depression. Imipramine was the first of the tricyclic antidepressants, so called because of their three-ring chemical bond, but others followed including amitriptyline (Elavil) in 1961. On this medication, depressed patients became more vivacious, had restored interest in their favorite activities, and experienced restful sleep and healthier appetites.

A third class of drugs known as the monoamine oxidase inhibitors (MAO-I) was developed in pharmaceutical laboratories. Nathan Kline pio-

neered phenelzine sulfate, marketed by such names as Nardil and Marplan. Its effectiveness was similar to other antidepressants but severe side effects were noted when patients took the medication in conjunction with foods containing tyamine, such as cheese, wine, and pickled foods.

The drug lithium, used during World War II as a substitute for table salt, was discovered to have antipsychotic properties by John Cade who worked in Australia. Although not approved by the FDA in the U.S. until the 1970s, it quickly became a standard treatment for bipolar disorder, or manic depression.

In the 1950s the first antianxiety drugs or minor tranquilizers were developed, the diphenylemythanes such as meprobamate (marketed under the trade name Miltown) was introduced in 1954. Later the benzodiazepine tranquilizers such as Librium or Valium were produced. These seemed to eliminate anxiety with few side effects, however the addictive properties of these drugs would receive a great deal of scrutiny in Congressional hearings and the popular press in the 1970s. In response, pharmaceutical companies sought less addictive versions of these drugs and produced Xanax in 1981, which was found to be effective for the newly named disorder panic attacks.

All of these drugs inspired interest in the neurotransmitters. First identified in 1920 by Otto Loewi, neurotransmitters are found in the synaptic gaps between the neurons of the brain. Although there are hundreds of them in the brain, research focused on two that seemed to be particularly important in both mood and behavior: dopamine and serotonin. Depressed patients were found to have less serotonin available than those who were not depressed. This led to the now prevalent belief that depression was caused not by environment but by changes in the brain chemistry such as a lack of free serotonin. Depression was therefore assumed to be as treatable as a biological disorder rather than to a psychological one. Researchers at Eli Lilly focused on creating a drug that would block the reuptake of serotonin into the neuron, thus was created an entirely new class of drugs—the selective serotonin reuptake inhibators (SSRIs). Lilly developed fluoxetine hydrochloride in 1974, but the drug was not approved by the FDA until 1987 when it was marketed under the name Prozac and became the established industry standard, perhaps the most well-known psychiatric drug of all time.

The psychopharmacological revolution that began with the accidental discovery of the clinical possibilities of a variety of drugs generated an interest in the specific effects of those drugs on brain chemistry, leading to the discovery of more specific treatment agents such as the SSRIs. For some this confirmed their belief that mental illness is biological rather than psychological.

See also **Medicine; Psychiatry, Diagnosis and Non-Drug Treatments**

TIMOTHY W. KNEELAND

Further Reading

American Psychiatric Association. *Essentials of Clinical Psychopharmacology*, Schatzberg, A.F. and Nemeroff, C.B. American Psychiatric Press, Washington, 2001.

Grob, G. *From Asylum to Community: Mental Health Policy in Modern America*. Princeton University Press, Princeton, 1991.

Healy, D. *The Creation of Psychopharmacology*. Harvard University Press, Cambridge, MA, 2002.

Shorter, E. *A History of Psychiatry from the Era of the Asylum to the Age of Prozac*. Wiley, New York, 1997.

Speaker, S. From "happiness pills" to "national nightmare": changing cultural assessment of minor tranquilizers in America, 1955–1980. *J. Hist. Med. Allied Sci.*, 52, 338–376, 1997.

Valenstein, E. Blaming the Brain: The Truth about Drugs and Mental Health. Free Press, New York, 1998.

Q

Quantum Electronic Devices

Quantum theory, developed during the 1920s to explain the behavior of atoms and the absorption and emission of light, is thought to apply to every kind of physical system, from individual elementary particles to macroscopic systems such as lasers. In lasers, stimulated transitions between discrete or quantized energy levels is a quantum electronic phenomena (discussed in the entry Lasers, Theory and Operation). Stimulated transitions are also the central phenomena in atomic clocks. Semiconductor devices such as the transistor also rely on the arrangement of quantum energy levels into a valence band and a conduction band separated by an energy gap, but advanced quantum semiconductor devices were not possible until advances in fabrication techniques such as molecular beam epitaxy (MBE) developed in the 1960s made it possible to grow extremely pure single crystal semiconductor structures one atomic layer at a time.

In most electronic devices and integrated circuits, quantum phenomena such as quantum tunneling and electron diffraction—where electrons behave not as particles but as waves—are of no significance, since the device is much larger than the wavelength of the electron (around 100 nanometers, where one nanometer is 10^{-9} meters or about 4 atoms wide). Since the early 1980s however, researchers have been aware that as the overall device size of field effect transistors decreased, small-scale quantum mechanical effects between components, plus the limitations of materials and fabrication techniques, would sooner or later inhibit further reduction in the size of conventional semiconductor transistors. Thus to produce devices on ever-smaller integrated circuits (down to 25 nanometers in length), conventional microelectronic devices would have to be replaced with new device concepts that take advantage of the quantum mechanical effects that dominate on the nanometer scale, rather than function in despite of them. Such solid state "nanoelectronics" offers the potential for increased speed and density of information processing, but mass fabrication on this small scale presented formidable challenges at the end of the twentieth century.

Discovery of Quantum Wave Effects in Electronics

The most important quantum effect used for electronic devices is the tunneling effect, the ability of electrons whose energy is not otherwise high enough to overcome or "tunnel" through an energy gap (arising for example from a junction with some other material). This energy gap creates a barrier that contains the electron's movement. Electrons (and holes) actually go through (rather than over) a potential energy barrier. The tunneling effect is associated with the wave characteristics of electrons and the Heisenberg Uncertainty Principle. This effect was first noticed in a theoretical calculation of tunneling between bound states in an atom by Friedrich Hund in 1927, who termed it "barrier penetration."

In 1928, quantum tunneling was shown by Ralph Howard Fowler and Lothar W. Nordheim to explain electron emission from a metallic surface under an intense applied electric field. The field lowered the potential barrier that normally prevented electrons from escaping the metal, allowing tunneling across this lowered barrier. In 1934 Clarence Zener modeled the electrical breakdown

of insulators (i.e. they began to conduct) with the idea of tunneling. The Zener effect in semiconductors occurs with application of a reverse voltage to a junction between *n*-type and *p*-type semiconductor material that is heavily doped. Beyond a certain breakdown voltage, large numbers of new carriers (electrons) "tunnel" across the junction to form the avalanche current that occurs at breakdown. In the mid-1950s after the development of techniques for growing heavily doped monocrystal germanium with precise geometries and controlled doping, the Zener effect was used to build Zener diodes with a predetermined "breakdown" voltage. The Zener diode could be used to stabilize and limit direct current (DC) voltages in circuits, and they became the earliest commercial quantum electronic devices.

After World War II, various striking new tunneling phenomena were discovered, and new devices based on tunneling were invented. In 1954 Pierre Aigrain made theoretical calculations of the tunneling effect in high-impurity germanium semiconductors. Building on these preparatory works, Leo Esaki invented the tunnel diode (also known as the Esaki diode) in 1957 while working for Sony Corporation. In a tunnel diode—a highly doped germanium *p–n* junction with a small depletion region—the *p*-material valence band and *n*-material conduction band nearly overlap. At low bias the tunneling of electrons and holes can occur. The diode has an unusual current–voltage characteristic curve as compared with an ordinary junction diode: current initially increases as the bias voltage is increased from zero, but when the voltage reaches a value comparable to the band gap of germanium, the tunneling current decreases. Therefore an anomalous negative resistance region appears, which allows oscillator action and the construction of high-frequency electronic oscillators and logic circuits based on tunnel diodes. The quantum-confinement effects of tunnel diodes, while having much increased circuit switching speeds and reduced power consumption, did not find widespread use due to difficulties in processing devices with heavy doping.

In 1962 Brian D. Josephson predicted the existence of a tunneling supercurrent that traversed a gap separating two superconductors. Ivar Giaever and others confirmed this superconducting tunnel effect experimentally using Josephson junctions consisting of superconducting thin films separated by a thin oxide barrier. In 1973, Josephson, Esaki, and Giaever were jointly awarded the Nobel Prize for Physics for the discovery of tunneling phenomena. Josephson junctions can be used for the construction of rapid single-flux quantum (RSFQ) electronics (proposed by Konstantin Likharev, Oleg Mukhanov, and Vasily Semenov in 1985). RSFQ electronics allows ultrafast circuitry with switching rates over 100 gigahertz (i.e., at least 300 times higher than the fastest similar semiconductor circuits) but are hampered by the necessity of their cooling to 4 to 5° Kelvin.

The concept of quantum confinement of electrons is used to trap electrons in specific locations in nanometer-scale structures. Quantum "islands" or "wells" between two closely spaced barriers (usually layers of two different III/V semiconductor alloys) are used in resonant tunneling diodes and transistors (invented by Esaki, R. Tsu, and L.L. Chang in 1974). Controlling the composition, shape, and size of the island permits control of the energy levels of the confined electrons such that electrons may pass through the device only by quantum tunneling. Resonant tunneling diodes can achieve very high switching speeds and have the speed performance of Josephson devices without the necessity of cryogenics. As MBE growth techniques improve in the early twenty-first century, chips incorporating resonant tunneling diodes and transistors will provide increased density and speed of information processing, consuming far less power than conventional integrated circuits.

The single-electron tunneling transistor, first proposed by Dmitri Averin and Konstantin Likharev in 1985, was developed by Theodore Fulton and Gerald Dolan at Bell Laboratories. Single-electron transistors, just a few nanometers wide, control single electrons by localizing them on nanoscale circuit elements linked by tunnel junctions. Such devices could represent bits of information with a single electron and be used in ultrahigh-density memory devices. However single-electron transistors have low voltage gain, are sensitive to random background charges, and until the late 1990s could not operate at room temperature.

The next generation of electronics will see advances in nanofabrication technology such as the ability of MBE to artificially engineer the band structure of a device and reliably produce nanometer-scale islands, barriers, and "heterojunctions" between islands and barriers. This will make it possible to constrain electrons so that they travel in a one-dimensional quantum wire only, or to create quantum dots ("zero-dimensional"), further increasing the performance and reducing the size of quantum electronic devices for integrated circuits.

See also **Clocks, Atomic; Josephson Junction Devices; Nanotechnology, Materials and Applications**

GILLIAN LINDSEY

Further Reading

Brown, L.M., Pais, A., Pippard, A.B. *Twentieth Century Physics*. American Institute of Physics Press, New York, 1995.

Hund, F. *The History of Quantum Theory*. Harrap, London, 1974.

Lundqvist, S., Ed. *Physics 1971–1980*. World Scientific, Singapore, 1992. Nobel lectures, including presentation speeches and laureates' biographies.

Nobel Foundation. *Physics 1963–1970*. Elsevier, Amsterdam, 1972. Nobel lectures, including presentation speeches and laureates' biographies

Turton, R. The Quantum Dot: A Journey into the Future of Microelectronics. Oxford University Press, New York, 1995.

R

Radar aboard Aircraft

As with much else in radar, airborne radar sets were first developed during World War II, and most of the modern uses for such sets were explored during that war. While airborne radar shares much in common with surface and naval sets, there are many factors involved that make airborne installations very different from either of the latter.

Generally speaking, the radar set itself needs to be made smaller to fit in the limited space available within an airframe. Antennas also need to be smaller; too large, and the drag on the aircraft will make it dangerous to fly. During World War II, a Luftwaffe directive hindered German airborne radar development by prohibiting external radar antennas on aircraft; it was rescinded in 1941. The concern was well founded as at least one test-bed aircraft crashed due to the performance problems caused by the external antennas. There are also only so many places on an airframe where a larger antenna can be placed, possibly affecting the usefulness of the radar. Another major consideration is power; an aircraft is limited in how much power it can produce, and it needs much of that power to fly safely. Additional power supplies can be added to the airframe, but these add to the weight of the aircraft. The size and weight of the set, the size and placement of the antenna(s), and the power consumed must all be balanced against the purpose of the radar set.

In the early days of World War II, the British used their ground-based radar network ("Chain Home") to direct fighters into position to attack German bombing raids. This was easy enough during the day and in good weather as the Chain Home sets had a margin of error of around 5 kilometers, which was close enough for the bombers to be seen by the fighter. At night, the visual range was reduced to around 300 meters. Something needed to bridge that gap if the fighters were to defend against night raids, so British scientists (under Sir Robert Watson-Watt) began work on an airborne intercept (AI) radar. AI needed to be at least accurate enough to get the fighters within that 300 meters. In order to increase the accuracy, AI radar needed to use centimetric radio waves (with a wavelength close to 10 centimeters) as opposed to the "long wave" (1- to 1.5-meter wavelengths) used by Chain Home.

Using a shorter wavelength helps to increase the accuracy of radar but causes other problems: all else being equal, a shorter wavelength fades more over a distance than a longer one (therefore needing more power in the transmission). The AI only needed a maximum range slightly exceeding the Chain Home margin of error, reducing power requirements, but existing wave generators still could not provide the necessary power at the shorter wavelength. Existing wave generators (klystron valves and magnetrons) either could not produce the centimetric waves needed or produced only very low-powered waves.

The necessary power at short wavelengths came in 1940 with the discovery of the resonant cavity magnetron by John Randall and Henry Boot at the University of Birmingham, U.K. Their invention produced a centimetric wave nearly 1000 times stronger than any previously produced. Like the magnetron, the cavity magnetron uses a magnetic field to produce oscillations in a stream of electrons moving through a vacuum. These oscillations are the source of the radio waves used by the radar.

However, the magnetron cannot efficiently produce wavelengths much shorter than 1 meter. The cavity magnetron adds a cylindrical metal shell with "resonant cavities" around its circumference. Under the influence of the magnetic field, several electron streams "pinwheel" around the central cathode. As the tips of the pinwheel pass over the resonant cavities, they produce oscillation in each cavity—the process has been compared to blowing into a whistle. This construction allows the cavity magnetron to oscillate much faster than the ordinary magnetron, resulting in shorter wavelengths.

In addition to greater accuracy, the shorter wavelengths also allowed antennas to be more compact, since antennas work best at a whole multiple of the wavelength being used. In aircraft, the antenna is most commonly a dipole one-half of the wavelength used. Early long-wave sets used antennas attached longitudinally to the sides of the fuselage, but these limited the radar to "seeing" only to the side. Forward-looking radar depended on long, cable-like antennas affixed along the leading edges of the wings.

Yagi antennas (named for a Japanese physicist who published papers about the antenna type in English) were also used in many designs. They resemble an old rooftop television antenna: a series of parallel rods in a plane, often tapering in length toward one end. German night fighters used four of these extending from the nose of the airplane, which was a crude method of electronically "steering" the radar beam. With centimetric radar, the Allies used a mechanically directed, parabolic "dish" antenna. Because they were using such short wavelengths, these antennas could be placed inside a fairing or within the nose of the aircraft.

Fast-acting duplexers were also important to AI development, as these allowed the radar set to transmit and receive using the same antenna. The duplexer is like a check valve that prevents the powerful transmitter energy from going directly to the sensitive receiver, which would overload it. Without a duplexer, a radar set would require two very accurately aligned antennas: one to transmit and one to receive.

Collaboration with American researchers at the Massachusetts Institute of Technology Radiation Laboratory (Rad Lab) resulted in AI sets operating at 10 centimeters with fine resolution and long range. Initially, these were installed in small bombers and large twin-engine fighters, but as the war progressed, the AI sets steadily got smaller and simpler to use, until they could be installed in single-seat fighters without greatly affecting their performance. German AI benefited from examining Allied radar sets in crashed airplanes and then producing some very good airborne radars. In general, the German effort lagged behind that of the Americans and British.

The centimetric AI sets were soon adapted for air-to-surface vessel (ASV) use. The Germans used long wave (1- to 1.5-meter) ASV radars that helped them locate the Allied convoys to Russia, but the primary Allied use of ASV was detecting submarines. The British had used long wave ASV since 1940, but those sets could only get the bombers within range of the submarine, which then had to be located visually, often giving alert U-boat crews a chance to crash-dive and get away. Also, once the Germans realized that the Allies were tracking their subs with radar, they began equipping the U-boats with "Naxos," a radar detector. Centimetric sets not only defeated the Naxos (it was not designed to detect radio emissions much shorter than 1 meter), but also increased resolution to the point where the periscopes and snorkels of submerged submarines could be picked out of the background clutter and the bomber could get close enough to drop depth bombs using ASV alone.

Each of the aforementioned radars used oscilloscopes to display information: range, height, and direction were often shown as spikes in a flat line. The next development was the British H_2S navigational radar, which used a plan position indicator (PPI). The PPI is the most familiar radar display: a circular screen with a line sweeping around it like the hand of a clock. The PPI shows graphically the horizontal relationship between the radar (at the center of the screen) and anything it detects. By connecting a PPI to the ASV radar, the ground clutter that made AI useless at too low an altitude instead made a picture in which water and roads showed up as dark areas and buildings and steep cliffs showed up brightly. By comparing the radar picture to a map, a radar operator could navigate by noting the location of prominent radar landmarks.

From 1946 to 1950, an American version of H_2S, designated AN/APS-10 was used in an experiment by American Airlines, whose pilots found it useful in avoiding dangerous weather, leading to the weather avoidance radar common to all modern commercial carrier aircraft. Commercial and general aviation aircraft also carry a radar altimeter that gives an accurate measurement of height above terrain that is very useful when landing using instruments alone.

One descendant of AI is the airborne early warning (AEW) aircraft. The most recognizable is

the U.S. E-3 Sentry, often called AWACS (for airborne warning and control system), which entered service in 1977. But the AEW idea goes back to the postwar period and is not limited to the U.S. Several other countries also build AEW airplanes and helicopters, and many more countries buy them. In the late 1980s the American military began developing the E-8 JSTARS (joint surveillance target attack radar system) aircraft. The E-8 is equipped with air-to-ground synthetic aperture radar that looks to the side and uses the motion of the airplane and lots of computing power to simulate a much larger antenna (the "synthetic aperture"). This produces radar images at a very high resolution, allowing the system to detect ground vehicles at a great distance.

At the close of the twentieth century, combat aircraft continued to use radar for the same purposes as in World War II: navigation, air and surface search, and targeting. Using computers and guided munitions, they could also automatically release bombs or launch missiles at the appropriate time. The big difference between 1945 and 2000 is that most, if not all, of these functions can be done by a single aircraft carrying a single radar with a range and resolution much greater than any airborne set used during the war.

See also **Radar, Defensive Systems in World War II; Radar, Displays; Radar, High Frequency and High Power**

LAURENCE BURKE

Further Reading

Brown, L. *A Radar History of World War II: Technical and Military Imperatives*. Institute of Physics Publishing, Bristol, 1999. A good history of radar developments in all countries during World War II.

Buderi, R. *The Invention that Changed the World: How a Small Group of Radar Pioneers Won the Second World War and Launched a Technological Revolution*. Simon & Schuster, New York, 1996. Concentrates on the MIT's Rad Lab and its people, but also includes good information on the British and Bell Labs efforts. Also follows many postwar "spinoffs" of wartime radar research.

Frieden, D.R., Ed. *Principles of Naval Weapons Systems*. Naval Institute Press, Annapolis, MD, 1985. Intended as a textbook for students at the U.S. Naval Academy, the descriptions of different systems (including radar) begin with basic information and quickly move toward equations that describe performance.

Jane's Avionics. Jane's Publishing, London, 2003. Published annually, this volume attempts to list every commercial avionic system (including radar) made in every country, whether civil or military.

Jane's Weapon Systems. Jane's Publishing, London, 1988. This volume includes military airborne radar systems, with a separate listing of air- and surface-launched missiles using radar guidance. Some time after 1988, *Jane's* began publishing separate volumes for air-launched weapons, missiles and rockets, naval weapons systems, electronic mission aircraft, and radar and electronic warfare systems.

Pritchard, D. *The Radar War: Germany's Pioneering Achievement 1904–1945*. Patrick Stephens, Wellingborough, 1989. Pritchard presents a comprehensive picture of German radar development, though he also gives a lot of technical details and model designations that might daunt the casual reader.

Rhodes, M. RADAR: wartime development—postwar application, AN/APS-10. *Aerospace Historian*, December, 231–240, 1981. Presents a brief history of the airborne search radar along with a description of its early postwar applications (including the American Airlines test).

Scott, P. Self-control in the skies: an onboard device for flying closer together more safely. *Scientific American*, 282, 1, 36, January, 2000. Using radar technology to prevent aircraft collisions.

Radar, Defensive Systems in World War II

With the onset of war in September 1939, Britain, Germany, and the U.S. had advanced radar designs while France, Russia, The Netherlands, Italy, and Japan had little of value in comparison although they had made research efforts along those lines. Of these endeavors, only Britain had proceeded past the prototype stage into a state of war readiness in the form of the Chain Home air defense. Germany had technically the best radar designs, but the Wehrmacht intended to wage a war of aggression and initially gave little support to a technology whose strength lay overwhelmingly in defense. In the U.S., because of the contentious battleship–bomber disputes of the 1920s, the Navy had pressed for any new technical method to defend ships against air attack, and the Army had sought to perfect its anti-aircraft artillery with methods of combating bombers at night.

Radar was not a factor during the opening months of the war except to provide Britain her narrow margin of victory in defeating the Luftwaffe's attempt to destroy the Royal Air Force. The Germans then resorted to night attacks on cities, intending to damage industrial production and intimidate the population. The Chain Home radars were ineffective for tracking planes over land, and the attacking formations met with impotent anti-aircraft artillery. This led to the introduction of new radar techniques in early 1941. A 1.5-meter radar formed a "searchlight" beam that could follow aircraft overland, allowing the bomber and the pursuing fighter to be tracked close enough for the fighter pilot to observe it,

visually at first and later with airborne radar. By May 1941 this method began to cause serious losses to the bombing fleets.

When the air war began to flow toward Europe, Germany built a system of radar-based fighter and anti-aircraft gun defenses that caused serious losses to RAF Bomber Command at night and the U.S. Army Air Forces during the day. In this they were greatly helped by the 50-centimeter equipment provided by the Telefunken Company for anti-aircraft guns and night fighters. Meterwave equipment with large arrays of dipoles were able to locate the attackers at ranges of 250 kilometers or more. From 1943 to 1945 the contending forces fought a highly technical electronic war in which new techniques were quickly met with countermeasures.

Britain developed airborne radar for use by the Fleet Air Arm that operated at 1.5-meter wavelength. This set, ASV Mk II, proved to be one of the most powerful air weapons in the war at sea. Mounted on the sturdy carrier biplane Swordfish, it located the battleship *Bismarck*, leading to her destruction. More importantly it located the Italian convoys carrying supplies to North Africa at night, resulting in their interdiction and in Rommel's defeat. The ASV Mk II, redesignated ASE for American service, became the U. S. Navy's eyes in the Pacific War when mounted on the long-range Catalina.

In September 1940 Britain sent a secret scientific technical mission to the U.S., called the Tizard Mission after its leader Henry Tizard. It opened Britain's technical war secrets, specifically radar, to the Americans on the assumption that the hoped-for future ally was lagging in this new field. It also opened up the vast American electronic industries for production, which was impossible if the technological details remained secret. American radar engineers found only two designs of interest: ASV and the resonant magnetron, a revolutionary new generator for waves of a few centimeters (microwaves). To exploit the latter a completely new research group was established at the Massachusetts Institute of Technology—the Radiation Laboratory.

A wavelength one-tenth of that of the 1-meter-wave sets previously in use allowed targets to be located on the surface of the earth, a difficult task for longer wavelength sets because of the confusion caused by surface reflections. This led to airborne sets that provided a map-like presentation of the ground, a development that was hoped might permit bombing German cities at night with some semblance of accuracy. Despite the technical sophistication of the equipment, this hope was not fulfilled, other than that the city was hit rather than the surrounding countryside. Much more useful was the installation of microwave sets for sea search aboard ships. The strong initial motivation for this was in the struggle with submarines, but it was soon found that the device was most useful in navigation. Ships could enter harbors in thick fog, hold their position in blacked-out convoys, and approach unmapped Pacific islands with greater safety.

Decimeter waves had been applied from the beginning of the war for directing fire against ships and aircraft, but microwaves improved the accuracy of such sets to a high degree. The most remarkable radar of this kind was the American SCR-584, a 10-centimeter, automatic tracking set intended primarily for directing anti-aircraft guns. This weapon was deployed in early summer 1944 in time to be used against the German V-1 flying bombs. When combined with an artillery fuse that sensed the presence of the target by its effect on a radio signal emitted by the fuse, anti-aircraft artillery became astonishingly effective. In the final days of the V-1 attacks, 95 percent of the bombs were brought down, mostly by anti-aircraft fire.

The entire nature of naval warfare in the Pacific was changed by radar. Most importantly, radar provided aircraft carriers with warning adequate to clear decks of planes, move bombs below, and flush petrol lines with carbon dioxide. When carriers were caught unaware with their thin decks over such explosive cargo, they simply blew up from a bomb that would not ordinarily have proved fatal. The Japanese fleet at Midway lost four carriers in this way. It was this extreme vulnerability that had given carriers a subservient role to heavily armored battleships in naval doctrine before radar's function became clear.

Japan did not deploy radar until mid-1942, generally in sets distinctly inferior to those of the Allies. However, Japan did invent the resonant magnetron independently, before the British in fact, and used it on occasion to the consternation of the Americans. Although the Soviet Union had also discovered the resonant magnetron in the late 1930s (dismissing it as of no importance and publishing details in open literature in 1940), it had only a few third-rate radars at the beginning of their struggle with Germany. Bombing of industrial targets by both sides did not figure in those gigantic land actions, so radar had little effect on the outcome. Germany used its superior radar to help maintain air superiority.

See also **Computers, Analog; Radar Aboard Aircraft; Radar, Long-Range Early Warning Systems; Radar, Origins to 1939**

LOUIS BROWN

Further Reading

Brown, L. *A Radar History of World War II*. Institute of Physics Publishing, Bristol, 1999.

Buderi, R. *The Invention that Changed the World*. Simon & Schuster, New York, 1996. A history of microwave radar.

Burns, R, Ed. *Radar Development to 1945*. Peter Peregrinus, London, 1988. A collection of 40 competently written articles.

Guerlac, H.E. *Radar in World War II*. Tomash, American Institute of Physics Publishers, New York, 1987. A history of the MIT Radiation Laboratory.

Howse, D. *Radar at Sea: The Royal Navy in World War 2*. Naval Institute Press, Annapolis, 1993.

Jones, R.V. *The Wizard War: British Scientific Intelligence, 1939–1945*. Coward, McCann & Geoghegan, New York, 1978. American edition of *Most Secret War*. Treats key elements of radar history.

von Kroge, H. *GEMA: Birthplace of German Radar and Sonar*. Institute of Physics Publishing, Bristol, 2000.

Lobanov, M.M. *Nachalo Sovetskoy Radiolokatsii* [The Beginnings of Soviet Radar]. Sovetskoye Radio, Moscow, 1975.

Nakagawa, Y. *Japanese Radar and Related Weapons of World War II*. Aegean Park Press, Laguna Hills, CA, 1997.

Price, A. *The History of US Electronic Warfare: The Years of Innovation—Beginnings to 1946*. Association of Old Crows, Washington D.C., 1984.

Reuter, F. *Funkmess: Die Entwicklung und Einsatz des RADAR-Verfahrens in Deutschland bis zum Ende des Zweiten Weltkrieges* [The development and deployment of radar in Germany to the end of World War II]. Westdeutscher Verlag GmbH, Opladen, 1971. Exhausts the material of the German archives.

Rowe, A.P. *One Story of Radar*. Cambridge University Press, Cambridge, 1948. Memoirs of the director of Air Ministry radar research.

Watson-Watt, R. *The Pulse of Radar: The Autobiography of Robert Watson-Watt*. Dial Press, New York, 1959. American edition of *Three Steps to Victory* by the inventor of Chain Home.

Radar, Displays

Those who first conceived of radar early in the century often envisioned systems that would simply indicate, perhaps by sounding a buzzer or lighting a lamp, that a target had been detected and where it was located. Those who first reduced radar to practice in the 1930s, however, were radio scientists who knew that the returning radar signals would somehow have to be distinguished against a background of radio-frequency interference and noise. They were accustomed to displaying signals visually on cathode ray tube (CRT) oscilloscopes, and they naturally turned to such means for radar. This made the operator an essential part of the radar system, responsible for the final stages of the detection process and extraction of target data.

The CRT was an invention of the late nineteenth century (Karl Ferdinand Braun, 1897) that would become familiar to people in the late twentieth century in the form of television display tubes and computer monitor screens (see Computer Displays). The CRT was typically housed in an evacuated glass envelope roughly conical, or funnel-shaped, in form. Near the apex or neck is the cathode, a source of electrons. By means of magnetic or electrostatic fields (either may be used, depending on application) the electrons are formed into a very narrow, tightly focused beam aimed toward the front or screen end of the tube, to which they are drawn by the anode. Additional sets of magnetic or electrostatic elements are provided to deflect the beam. Since the mass of the beam is very small and the forces relatively large, it can be moved from spot to spot anywhere on the screen at a rate fast enough to seem instantaneous to a human observer. The point at which the beam intersects the screen is made visible by coating the inside of the CRT face with a phosphor that emits photons on absorbing electrons.

The basic radar display was what came to be called an A-scope, which provided a plot of signal amplitude against range. On the basis of signal strength (resulting in a peak rising above the grass) and width, operators learned to distinguish radar returns from the "grass" (so called from its appearance on the A-scope display, see Figure 1)

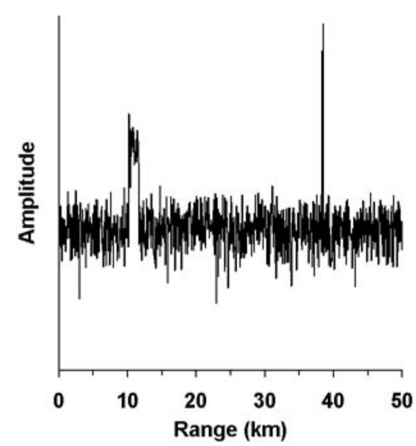


Figure 1. Typical A-scope trace showing noise or "grass" at all ranges, clutter return between ranges of 10.200 meters and 11,600 meters, and strong aircraft target return at range of 38,300 meters.

resulting from noise and from the "clutter" of interference and unwanted or spurious targets. The A-scope provided an immediate measurement of range, and the azimuth or elevation could be determined by noting the position of the antenna at the time of detection.

Although such a system worked, it was of course cumbersome and slow. The 1940 British invention of the plan position indicator (PPI) provided a far more practical display for search and general orientation purposes. On a PPI display, range was proportional to distance from the center point, azimuth was directly translated to azimuth from an index on the CRT face, and signal strength was displayed by modulation of the intensity of the beam. By reducing sensitivity or gain to a level that caused most of the noise background to disappear, targets could be seen as bright spots. Range and azimuth could be read directly from the display; and by marking the point on the screen at which the target appeared during successive sweeps of the radar antenna and connecting the dots using grease pencil, the target's track relative to the radar position could be displayed directly on the face of the PPI. This greatly improved the efficiency of radar as an aid to directing fighters to intercept bombers, as well as for navigating ships in order to approach targets and avoid collisions. The A-scope and a variety of other specialized displays remained valuable for particular uses, however. A-scopes gave more precise range determination and discrimination between targets and clutter, for instance.

Radar displays improved as a result of development of better focusing and deflection systems for CRTs as well as superior phosphors. More fundamental, however, were improvements in the electronic processing of the signals that were input to the display. Physical considerations as well as careful measurements provided information about electronic characteristics that could be used to distinguish wanted signals from unwanted noise and clutter. A simple example is the case in which the target of interest is expected to move toward or away from the radar at a velocity that differs from that of clutter and interfering sources. If the radar is able to measure the Doppler frequency shift—proportional to velocity—then processing can suppress returns with the "wrong" shift values, resulting in great improvements in the ratio of signal to interference and noise ($S/(I+N)$) on the display. Improvements in knowledge of signal, interference, and noise characteristics as well as advances in electronics brought about progress in displays.

From 1975, digital technology had an increasing role in radar displays. As digital circuit speeds increased, approaching and even exceeding the radio frequencies of radars, it became possible to digitize more and more of the radar's processing chain, allowing for far more complex computations. In many cases, the radar's computer calculated the most likely estimated target positions and characteristics and presented a "synthetic" picture of this on a display that was really a computer monitor. When noise and interference levels were especially high, the computer might be instructed to try to form coherent tracks of successive observations before displaying anything at all, so that the first thing the operator saw was a track already extending seconds or even minutes into the past.

Owing to the increasing dominance of computers, displays largely disappeared as a separate topic in texts on radar after about 1980. Thus over the course of the century, radar displays in a sense came full circle, returning to the original concept of a system that simply reported the desired target data and left all else aside. Where maximum discrimination is needed, there often remained a role for operators in examining and evaluating data, but only after the data had first been subjected to intensive computer processing. As in many other fields of technology, by the end of the twentieth century, computers usually stood between humans and their radars.

See also **Computer Displays; Radar, Origins to 1939**

WILLIAM D. O'NEIL

Further reading

Berg, A.A. Radar indicators and displays, in *Radar Handbook*, 1st edn., Skolnik, M.I., Ed. McGraw-Hill, New York, 1970.

Ridenour, L.N., *et al*. The gathering and presentation of radar data, in *Radar System Engineering*, Louis N. Ridenour, Ed. McGraw-Hill, New York, 1947.

Skolnik, M.I. *Introduction to Radar Systems*, 3rd edn. McGraw-Hill, New York, 2000.

Radar, High Frequency and High Power

High-Power Microwave Radar

While early radar designers were driven to frequencies of more than 1000 megahertz by considerations of the availability of high-power components, it was appreciated very early on that higher frequencies and thus shorter wavelengths would allow better precision. Frequency and

wavelength are inversely related according to the equation

$$\text{Wavelength} = c/\text{frequency}$$

where c = velocity of light = 3×10^8 meters per second.

The need for higher frequencies became acute with sets carried by aircraft, where practical antenna dimensions gave very inadequate accuracies at the only available (i.e., low) frequencies that were possible until 1940.

This need was met through one of the most important of radar technology breakthroughs, the 1939 U.K. development of a high-power version of the earlier magnetron. This "cavity magnetron" was a thermionic valve (vacuum tube) in which an arrangement of tuned cavities in the anode structure modulated electron motion through their effect on magnetic field geometry. It was quickly applied to radars operating at frequencies of around 3 gigahertz, rising to around 9 gigahertz later in World War II. Because the wavelengths of these radars were so much shorter than those of earlier sets (approximately 3 to 10 centimeters versus 1 meter), they were termed "microwave" radars. The compactness and accuracy of powerful microwave radars, an achievement not duplicated by the German and Japanese enemies, played a significant role in the Allied victory. By the end of the war, peak powers of more than 1 megawatt were available at a frequency of 3 gigahertz.

Microwave radar became the dominant type for the majority of applications after 1945, employing a variety of forms of power tubes. Microwave frequencies generally were not used, however, in applications requiring very long range where the higher power and lesser atmospheric attenuation available at lower frequencies made them preferable. Antennas were often space-fed parabolic reflectors, producing a flat wave front from a single point source of radiation.

An important post-World War II development was the "monopulse" technique for precise tracking. This normally involved the use of multiple feeds with a common reflector to form several separate beams looking in slightly different directions. By comparing the amplitudes (or occasionally the phases) of the signals on different beams, the radar was able to calculate the precise azimuth and elevation of the target on the basis of a single pulse detection.

From the 1950s, phase-coherent microwave radars were progressively developed to permit detection and tracking of small moving targets against strong backgrounds of stationary clutter on the basis of the Doppler shift of the frequency of the received signal as compared with that transmitted. The first coherent radars were generally at lower frequencies and used continuous-wave rather than pulsed signals. As technology developed, the technique was extended to microwave and pulsed radars. Depending on the pulse repetition frequency employed, either the range or Doppler velocity information would be ambiguous; the choice would thus depend on the application's needs. By the 1960s, pulse Doppler microwave radars were beginning to be used in airborne service.

The 1970s brought the first phased-array microwave radars, allowing a single antenna to look in multiple directions simultaneously or in very quick succession. This involved hundreds or thousands of separate antenna elements, each separately fed by a signal whose phase could be independently varied in a systematic manner. Phased arrays were particularly attractive in applications where a computer steered the beams to track many targets at once. By the 1980s, the power outputs possible with microwave solid-state devices began to make it feasible to consider them as transmitter elements for phased arrays in place of elements fed from a central power tube through intervening phase-shifters. Even at the end of the twentieth century, however, microwave solid-state radars were in limited use owing to high costs of early generation components.

The 1990s saw the introduction of microwave active electronically scanned array (AESA) radars. In these, each transmitter element was coupled on the same electronic module or even the same chip, with a complete receiving element that detected and digitized its portion of the incoming signal, thus allowing great flexibility and precision of operation. Again, component costs slowed its acceptance.

High-Frequency Over-the-Horizon Radar

Radars operating in the high-frequency (HF) band (3 to 30 megahertz) may detect targets well beyond the nominal horizon through two mechanisms: "sky wave" and "surface wave." Early in the century, it was discovered that high-frequency radio waves were strongly refracted by the ionosphere. A HF beam aimed near the horizon would, under suitable conditions, be effectively reflected, returning to sea level some hundreds to thousands of kilometers from its transmission site. From the 1940s, interest developed in using this sky-wave transmission phenomenon to provide surveillance

at great ranges. Early HF over-the-horizon radars (OTHRs) were bistatic "forward scatter" systems in which a widely separated transmitter and receiver detected and tracked targets lying between them. Ballistic missile tracking was a major application.

By the 1980s technology had advanced sufficiently to support development of "backscatter" OTHRs having transmitter and receiver relatively close together to provide surveillance at distance. Typical transmitter arrays covered a line hundreds of meters long and broadcast at an average power of hundreds of megawatts. Receive arrays generally were more than 1 kilometer in length. Using Doppler processing to distinguish moving targets from stationary clutter, OTHRs searched annular sectors typically around 60 degrees in width, with a depth of several hundred kilometers, at ranges which varied (with frequency and ionospheric conditions) from about 500 to 3500 kilometers. Spatial accuracy was coarse, but Doppler velocities could be distinguished finely. Propagation of signals was affected by ionospheric conditions and could be disrupted completely at some periods. Reliability of operation was a function of location and orientation, being worst for systems in high latitudes looking toward the pole, where auroral phenomena affect the ionosphere (see Figure 2).

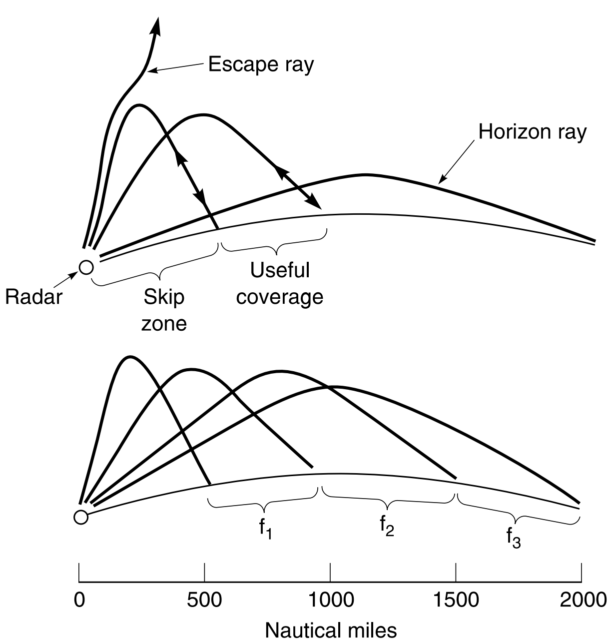


Figure 2. Vertical paths for sky-wave over-the-horizon radar (OTHR). Different frequencies, f_1, f_2, and f_3, cover different range intervals. (Ranges shown in nautical miles = 1852 meters; vertical scale exaggerated for clarity.)

Surface-wave high-frequency OTHRs saw less use. They depended on the diffraction of vertically polarized HF waves around the curve of the earth, especially over the smooth conductive surface of the sea. With high powers and Doppler processing, such radars could detect sea-level targets at ranges up to several hundred kilometers. OTHR applications included surveillance of aircraft, missiles, and ships for military purposes and control of smuggling as well as measurement of phenomena associated with ocean waves.

See also **Radar Aboard Aircraft; Radar, Origins to 1939; Radar, Defensive Systems in World War II**

WILLIAM D. O'NEIL

Further Reading

Boyle, D. OTH radar clinging to the surface. *Int. Defense Rev.*, 4, 501–503, 1989.
Brookner, E., Ed. *Radar Technology*. Artech, Dedham, MA, 1977.
Ferraro, E. and Ganter, D. Cold War to counter drug. *Microwave J.*, March, 82–92, 1998.
Foxwell, D. Beyond the horizon, but not out of sight. *Jane's Int. Defense Rev.*, 8, 48–51, 1997.
Guerlac, H.E. Radar in World War II, in *History of Modern Physics*, vol. 8. Tomash, 1987.
Headrick, J.M. HF over-the-horizon radar, in *Radar Handbook*, 2nd edn., Skolnik, M.I., Ed. McGraw-Hill, New York, 1990.
Headrick, J.M. Looking over the horizon. *IEEE Spectrum*, July, 36–39, 1990.
Headrick, J.M. and Thomason, J.F. Applications of high frequency radar. *Radio Science*, 33, 1045–1054, 1998.
Klass, P.J. HF radar detects Soviet ICBMs. *Aviation Week and Space Technology*, December 6, 38–40, 1971.
Kolosov, A.A., *et al*. *Over-the-Horizon Radar*. Artech House, Boston, 1987.
Ridenour, L.N., Ed. *Radar System Engineering*. McGraw-Hill, New York, 1947.
Skolnik, M.I. *Introduction to Radar Systems*, 3rd edn. McGraw-Hill, New York, 2000.

Radar, Long-Range Early Warning Systems

During the 1930s, Great Britain was one of several countries, including most notably Germany and the U.S. that experimented with radar for early warning of air attacks. The British "Chain Home" system, designed by Sir Robert Watson-Watt and established by 1939, included a string of stations along the east and south coasts. By mid-1940, most of the stations featured two 73-meter wooden towers, one holding fixed transmitter aerials and the other receivers. When it was discovered that low-flying aircraft could slip undetected beneath the original fence, Britain created a second string of "Chain Home Low" stations, beginning with

Truleigh Hill. The latter sites consisted of two separate aerials, one to transmit and the other to receive, mounted on 6-meter-high gantries and short enough to allow an operator inside the equipment hut beneath the gantry to manually rotate the arrays. Together, Chain Home and Chain Home Low provided a detection range of 40 to 190 kilometers depending on an incoming aircraft's altitude. This early warning capability contributed immeasurably to the RAF victory over the Luftwaffe in the Battle of Britain.

American forces failed to make equally beneficial use of radar to warn of the Japanese attack on Pearl Harbor on December 7, 1941. During late 1941, U.S. Army personnel in Hawaii had been field-testing SCR-270-B mobile units capable of detecting aircraft more than 160 kilometers away. Unfortunately, the system was undermanned by relatively inexperienced people. An hour prior to the December 7 attack, three of the five operational radar sites—Kaaawa, Opana, and Kawailoa—saw large reflections approximately 220 kilometers north of Oahu. All locations reported these sightings to the Information Center, where plotters put the data on a master board. Nobody at the Information Center interpreted the plots as unusual. The Opana unit continued to track the incoming target for more than 30 minutes, until ground clutter interfered with the signal just 32 kilometers from the coast. By that time, no one was on duty at the Information Center to plot the data. Not until several days later did the Americans realize that their radar stations had accurately detected the approach of the Japanese aerial armada, which might have enabled the Army, Air Force, and Navy aircraft to respond earlier and thereby limit the disaster.

Germany, Japan, and the Soviet Union also employed early warning radar during World War II with varying degrees of success. A network of Freya radars known as the "Kammhuber Line" enabled the Germans to detect approaching British and American bombers at a range of 95 kilometers and, when augmented by Würzburg radars, proved especially effective in determining exact range and height as the aircraft came closer. Japan relied on Mark-1 and Tachi-6 units—with ranges of 120 and 200 kilometers, respectively—to alert the home islands of impending B-29 raids late in the war. In early 1942, the Soviet Union installed RUS-2 radars, with a range of 95 to 145 kilometers, to aid locally in the defense of Moscow and Leningrad. While these various radar applications for early warning probably had little influence on the course of World War II, they were seedlings from which greater capabilities would spring.

The use of long-range, early-warning radar equipment for air defense grew from these roots to blossom into complex, integrated systems during the Cold War era. Since the most direct route by which Soviet aircraft could attack North America was across the polar cap, the U.S. and Canada cooperated during the 1950s to orient a multilayered early warning capability in that direction. In 1951, they began construction of the CADIN/Pinetree line, a series of more than 30 stations situated approximately along the U.S.–Canada border. To this was added the distant early warning (DEW) line in 1956—a string of as many as 70 sites stretching along the Arctic Circle from Alaska to Greenland. In 1958 the mid-Canada line—a Doppler electronic fence containing as many as eight sector stations and 90 unmanned sites along the fifty-fifth parallel—completed the original system (Figure 3). Upgraded over the years, the entire system was eventually replaced in 1985 by the north warning system, which was comprised of 13 long-range sites and 39 shorter-range stations. In 2001, the U.S. Air Force contracted with Lockheed Martin Corporation to upgrade operating software for long-range, atmospheric early warning radar systems at 33 locations by August 2006.

Concerned that the U.S. needed even earlier detection of incoming Soviet bombers or low-flying cruise missiles, the U.S. Air Force had contracted with General Electric Aerospace to begin building a prototype over-the-horizon-backscatter (OTH-B) radar in Maine during the mid-1970s. That system, designated AN/FPS-118, achieved limited operational capability in 1988. By bouncing signals off the ionosphere, the powerful OTH-B radar could locate and track targets thousands of miles from U.S. air space. The U.S. Air Force accepted both the East Coast and West Coast OTH-B systems from the contractor in 1990, but the end of the Cold War and a desire on the part of Congress to cut operating costs soon led to shutdown of the western portion. In 1994, Congress directed that the eastern system be used for counter-narcotics detection as well as for monitoring weather and the ocean environment. Three years later, that system was also mothballed.

To counter the threat of surprise intercontinental ballistic missile (ICBM) attacks by the Soviet Union, the U.S.—as a matter of the highest national priority—began construction of the ballistic missile early warning system (BMEWS) in 1959. Otherwise known as the 474L system, it

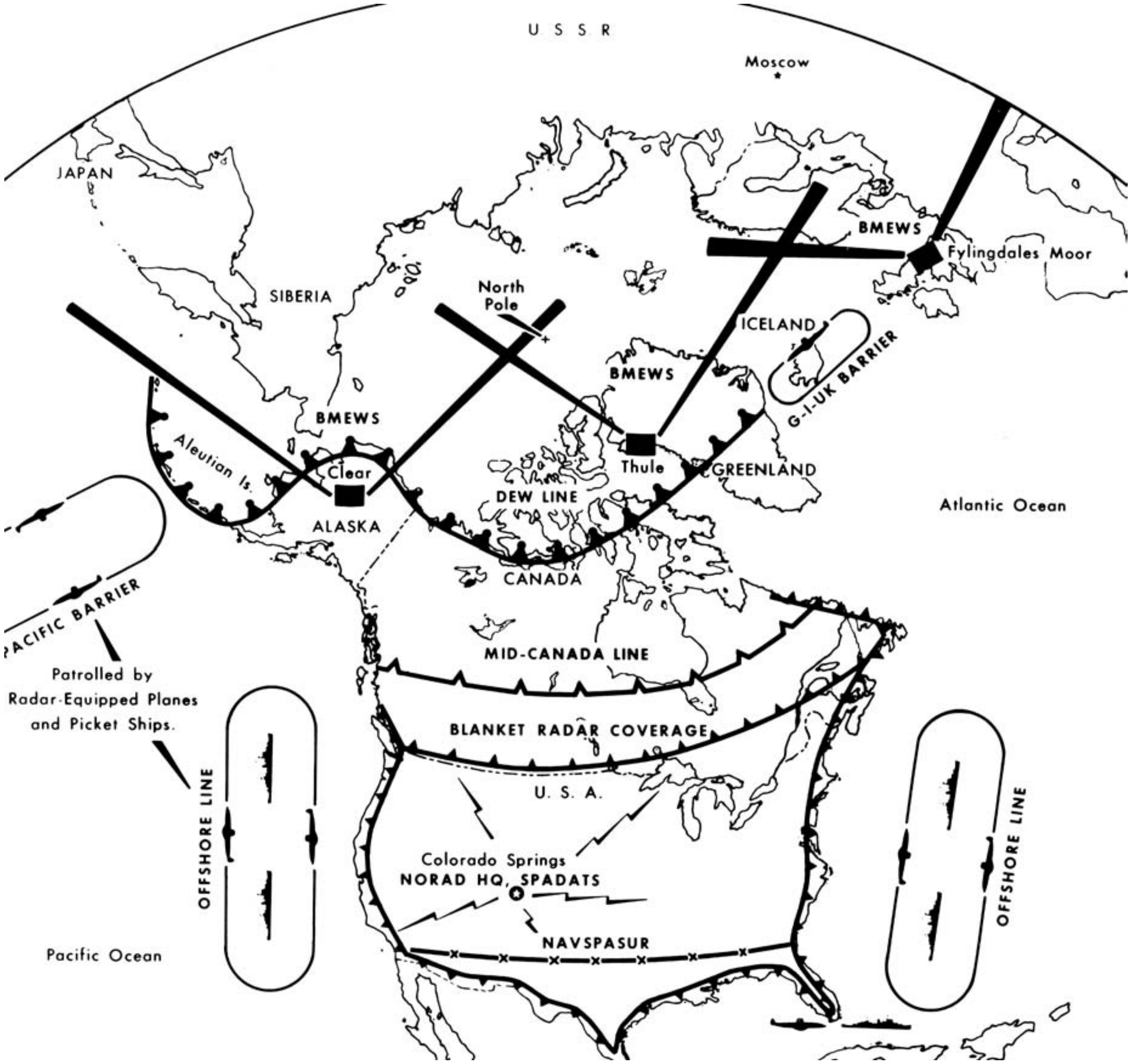


Figure 3. Location of U.S. early warning sensors, circa 1962. [*Source: Courtesy of U.S. Air Force.*]

included three sites: Thule Air Base, Greenland (1960); Clear Air Force Station (AFS), Alaska (1961); and RAF Fylingdales Moor, England (1963). The original BMEWS equipment consisted of two principal items: the AN/FPS-50 detection radar, which included three fixed antennas, each 50 meters high and 120 meters long; and the AN/FPS-92 tracking radar, a mechanical antenna 25 meters in diameter. During the 1970s, the increased threat of a massive sea-launched ballistic missile (SLBM) attack caused the U.S. to supplement BMEWS coverage with a long-range, solid-state phased-array radar (SSPAR) system called PAVE PAWS. Two of the latter, designated AN/FPS-115 and possessing a detection range of 3000 nautical miles, became operational in 1980 at Otis Air Force Base (AFB) in Massachusetts and Beale AFB in California. Others were built later at Robins AFB in Georgia and Eldorado AFS in Texas, but they were shut down in the 1990s as a cost-saving measure. Between 1985 and 2001, the U.S. also replaced the original BMEWS equipment with SSPAR technology.

During the Cold War, the USSR maintained an even more extensive early warning system for both ballistic missile and air defense. By 1985, the Soviets had deployed over 7000 air surveillance radars of various types, including over-the-horizon

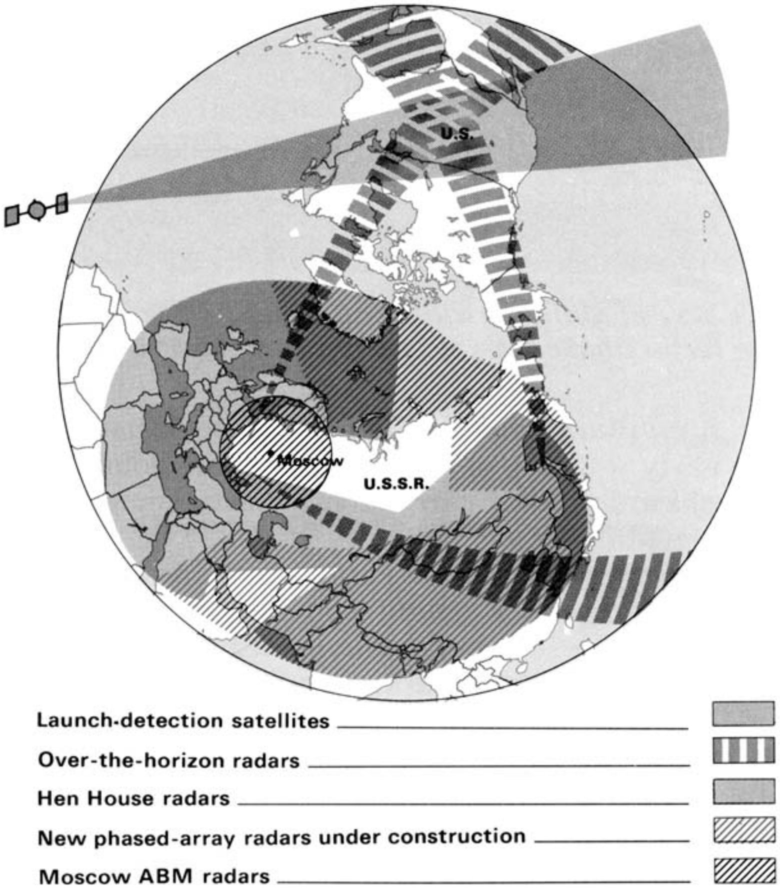


Figure 4. Location of USSR ballistic missile and tracking systems in 1985. [*Source: Courtesy of the U.S. Department of Defense.*]

models, at approximately 1200 locations, which gave them coverage at medium to high altitudes over all their territory and in some areas for hundreds of kilometers beyond their borders (Figure 4). They were also deploying new systems for early warning of cruise missile and bomber attacks. For ballistic missile warning, the USSR operated eleven large "Hen House" detection and tracking radars at six peripheral locations, as well as new phased array equipment. With the USSR's collapse in the early 1990s, however, the missile warning radar system became too fiscally burdensome. In 2001, Russia announced its intention to scrap many of the fixed installations and rely on mobile stations.

See also **Radar, Defensive Systems in World War II; Radar, High Frequency and High Power; Radar, Origins to 1939**

RICK W. STURDEVANT

Further Reading

Brown, L. *A Radar History of World War II: Technical and Military Imperatives*. Institute of Physics Publishing, Bristol, 1999.

Clark, D.L. Early advances in radar technology for aircraft detection. *Lincoln Lab. J.*, 12, 167–179, 2000.

Guerlac, H.E., Ed. *Radar in World War II*, 2 vols. Tomash, American Institute of Physics Publishers, New York, 1987.

Jockel, J.T. *No Boundaries Upstairs: Canada, the United States, and the Origins of North American Air Defence*, 1945–1958. University of British Columbia Press, Vancouver, 1987.

Nakagawa, Y. *Japanese Radar and Related Weapons of World War II*. Aegean Park Press, Laguna Hills, CA, 1998.

Schaffel, K. *The Emerging Shield: The Air Force and the Evolution of Continental Air Defense, 1945–1960*. Office of Air Force History, Washington D.C., 1991.

Stone, M.L. and Banner, G.P. Radars for the detection and tracking of ballistic missiles, satellites, and planets. *Lincoln Lab. J.*, 12 217–243, 2000.

U.S Department of Defense. *Soviet Military Power 1985*. Government Printing Office, Washington D.C., 1985.

Winkler, D.F. *Searching the Skies: The Legacy of the United States Cold War Defense Radar Program*. U.S. Air Force Air Combat Command, Langley AFB, June 1997.

Useful Websites

Belmont, L. *Radar in WWII*: http://www.strandlab.com/radar/

Canadian Minister of Public Works and Government Services. *Post World War II Radar*: http://www.dnd.ca/commelec/brhistory/anxf_e.htm
Federation of American Scientists, AN/FPS-115 PAVE PAWS Radar: http://www.fas.org/spp/military/program/track/pavepaws.htm
McManus, G.P. *BMEWS—510 Full Days*: http://www.bwcinet.com/thule/index.html
Toon, T.H. *World War Two—The Chain Home Low Radar*: http://schoolsite.edex.net.uk/468/radar.htm

Radar, Origins to 1939

Reflection was an important part of Heinrich Hertz's 1887 demonstration of the existence of electromagnetic waves, and the idea of using that property to "see" in darkness or fog was developed shortly afterwards.

Christian Hülsmeyer constructed a device in 1902 that he hoped might prevent collisions at sea. It used two cylindrical paraboloid reflectors to transmit and receive waves of decimeter length generated from spark oscillations with the reflected signals detected by coherer. The equipment was demonstrated successfully at a conference of ship owners in Rotterdam, but it was only capable of showing the direction of an object and not its range, as this required accurate timing of signals at the microsecond level, a technique that lay years in the future. Thus a liner eight kilometers distant was indistinguishable from a tug at 500 meters, and the remarkable device aroused no enthusiasm among seamen. Hülsmeyer's set was soon forgotten, although it was adequately patented and demonstrated to numerous witnesses. The idea recurred to no effect in World War I but seems to have been discussed informally among engineers. Guglielmo Marconi proposed using reflected radio waves for the location of objects in a paper delivered at a meeting of the Institute of Radio Engineers in New York in 1922.

By 1920 the vacuum triode had revolutionized the generation and detection of radio waves, and broadcast radio was transforming many aspects of everyday life. However the circuit elements needed to measure the time between the emission and reception of a pulsed train of waves—the key to radar ranging—were still missing. Triodes were able to work at high frequencies, and in the early 1930s various experimenters built equipment capable of measuring the speed of an automobile, if not its range. If the target were moving, the reflected wave would have its frequency shifted by the Doppler effect and through interference with the transmitted signal within the receiver would produce an easily recognizable signal, the beat frequency of which was proportional to the speed. Police radar was therefore possible before air-warning radar. A revival of the Hülsmeyer idea returned in 1935 in equipment designed by Camille Gutton and mounted in the new transatlantic liner *Normandie*. The equipment did not do well at sea, and the watch officers were not impressed because there was still no range information.

Triodes were unable to follow the rapid changes in signal amplitude that characterize a short wave train. Furthermore, operators required a device that presented the time elapsed between emission and reception. It was recognized that a cathode-ray oscillograph would perform this function, but those available before 1930 were inadequate for a variety of reasons. Both of these functions were also vital for television of sufficiently high definition to rival the cinema, and both were the subject of research in the electronic industry, which demonstrated all-electronic television in 1930. The necessary elements were multigrid amplifier valves and high-vacuum, low-voltage cathode-ray tubes. When these two circuit elements became available for the video amplifier and for the picture tube, respectively, serious radar work could begin.

The radar sets first envisioned were to use wavelengths of a few centimeters, which allowed the beam to be shaped into a form of "radio searchlight" with a reflecting dish of practical size. This approach faced a serious obstacle: the absence of any generator working at these short wavelengths with sufficient power or frequency stability. There were, however, many observations (through the Doppler effect) of aircraft and ships seen at wavelengths of a few meters, an effect that was particularly pronounced in experiments studying the propagation of waves intended for the transmission of television. At these frequencies, antennas of manageable size could be fashioned from arrays of dipole radiators so that the individual radiations would constructively interfere to form the desired radio searchlight.

By the early 1930s, serious efforts were underway in the U.S., Germany, and Britain to construct radio-location devices using relatively long wavelengths. (Russian efforts were ahead in the early 1930s, but they yielded little as a result of serious organizational problems and purges that sent key engineers to the gulag.) The German company GEMA built the first device that can be called a functioning radar set in 1935 with Britain and America following only months behind. Two groups in the U.S.—the Signal Corps and the Naval Research Laboratories—proceeded independently but on lines very similar to those of the

Germans in using dipole arrays. They had air-warning and searchlight-pointing prototype sets ready for production in 1939.

The British physicists Robert Watson Watt and Arnold Wilkins proceeded along a different line using wavelengths of tens of meters with broadcast rather than "searchlight" transmission. This equipment, although inferior to that working on shorter wavelengths, was seen by Air Vice-Marshal Hugh Dowding as the key to the air defense of Britain from expected German attack. As commander of the newly created Fighter Command, he created a system of radar stations and ground observers linked by secure telephone lines to the fighter units. He drilled Fighter Command to use the new technique, and when the Luftwaffe came in the summer of 1940, the attacking squadrons were ambushed by defending fighters positioned by radar.

Triodes capable of generating significant amounts of power at decimeter wavelengths had been developed by the late 1930s, and by 1940,the Bell Laboratories in the U.S., the Royal Navy in the U.K., and the Telefunken Company in Germany had sets that worked at 50 centimeters.

All these designs went into production with the onset of World War II and furnished, with various modifications, the radar used by the combatants until 1943, when centimetric-wave equipment was developed, which used electronics of a completely different nature.

See also **Radar, Defensive Systems in World War II; Radar, Long-Range Early Warning Systems**

LOUIS BROWN

Further Reading

Allison, D.K. *New Eye for the Navy: The Origin of Radar at the Naval Research Laboratory*. NRL Report 8466, U.S. Government Printing Office, Washington D.C., 1981.

Brown, L. *A Radar History of World War II: Technical and Military Imperatives*. Institute of Physics Publishing, Bristol, 1999.

Burns, R.W. *Television: An International History of the Formative Years*. Institution of Electrical Engineers, London, 1998.

Guerlac, H. The radio background of radar. *J. Franklin Inst*., 250, 285–308, 1950.

Jones, L.F. A study of the propagation of wavelengths between three and eight meters. *Proc. Inst. Radio Eng*., 21, 349–386, 1933. The remarkable behavior of waves later to be used for radar.

von Kroge, H. *GEMA: Birthplace of German Radar and Sonar*. Institute of Physics Publishing, Bristol, 2000.

Nakagawa, Y. *Japanese Radar and Related Weapons of World War II*. Aegean Park Press, Laguna Hills, CA, 1997.

Swords, S.S. *Technical History of the Beginnings of Radar*. Peter Peregrinus, London, 1986.

Taylor, A.H. *Radio Reminiscences: A Half Century*. U.S. Naval Research Laboratory, Washington D.C., 1948, 1960.

Watson-Watt, R. *The Pulse of Radar: The Autobiography of Robert Watson-Watt*. New York: Dial Press, 1959. American edition of Three Steps to Victory.

Radio: AM, FM, Analog, Digital

The term "radio" includes many different modes of wireless transmission. In the late nineteenth and early twentieth century, wireless telegraphy used spark-gap technology and intermittent waves to transmit Morse signals. Only with the development of continuous wave transmission in the early twentieth century did wireless telephony become possible, allowing effective transmission of the human voice and music. All wireless or radio signals consist of a carrier wave (a continuous wave, altered by amplitude or frequency, to which is attached the intended voice or music intelligence being transmitted) and one or more sidebands (the band of frequencies produced by modulation). The intended information or content of a given signal is carried on one of the modulated sidebands. While only analog amplitude modulation (AM) was used until about 1940, frequency modulation (FM) became increasingly important after that date. Digital transmission developed beginning in the 1980s.

AM

Amplitude modulation indicates that the strength of the sideband signal is modulated (several thousand times per second) in accordance with the amplitude or strength of the carrier wave. Broadcast radio, which was developed in the 1920s, was assigned to medium-wave frequencies in most nations. In the U.S., AM radio stations are assigned to 10-kilohertz channels; in much of the rest of the world, AM or medium-wave radio uses 9-kilohertz channels. While AM stations use narrower channels (and thus less frequency space) than FM stations, AM is prone to natural and most man-made electrical interference, or noise, which cannot be separated from the desired signal. Attempts to overcome static using more transmitter power failed. In part due to their narrow bandwidth and crowding on the AM band, signal response or "sound quality" is much poorer for AM (up to about 5 cycles per second) than FM, which regularly provides a signal response of up to 15 cycles per second due to its greater bandwidth.

Because they are located on medium-wave frequencies in most nations (540 to 1705 kilohertz in the U.S.), AM broadcast stations utilize ground wave propagation during daylight hours and sky wave propagation at night. Sky wave propagation, in which signals are bounced back to earth from the ionosphere, can carry a signal hundreds of kilometers, especially on cold, clear nights, though not in a predictable fashion. Station coverage or ''reach'' therefore varies by time of day and season.

FM

Frequency modulation signals vary by a swing of frequency rather than power output within the assigned channel. Edwin Howard Armstrong found that using a channel 20 times wider than an AM channel (200 kilohertz) would allow an analog signal with excellent frequency response (up to 15,000 cycles per second) that could avoid atmospheric interference (e.g., static from electrical storms) and much man-made interference as well. An FM signal needs to be only twice as strong as a more distant competing transmitter to suppress the interfering signal. Utilizing very high-frequency (VHF, or VHF radio outside of the U.S.), FM signals are propagated by direct line-of-sight means day or night, limiting transmitter coverage to a radius of no more than 95 to 110 kilometers depending on local terrain, but eliminating multipath AM interference.

FM radio was developed from 1928 to 1933 by Edwin Armstrong, a professor at Columbia University in New York and prolific radio inventor. Armstrong fought many patent battles, as did most early American radio inventors. Two were very important and lasted for years: the fight with inventor Lee de Forest from 1914 to a Supreme Court decision two decades later over the rights to the regenerative circuit, which he lost; and the battle over his basic FM patents, fought against RCA and only settled after his death in 1954. After considerable experimentation by Armstrong and others, FM was introduced as a broadcasting service in the U.S. in 1941, on 42–50 megahertz. About 50 stations were on the air before a wartime freeze was imposed on new construction in 1942. Television standards approved in 1941 required FM for the sound portion of the signal. After extensive research and hearings, the Federal Communications Commission (FCC) in early 1945 reallocated FM to its present 88–108 megahertz range, thus providing more channels but at a cost to the medium of ''starting over'' in the face of television and revived AM competition.

In 1955 the FCC allowed FM stations to multiplex an additional nonbroadcast signal as a means of generating revenue. The service usually provided background music to offices and stores. Six years later, stereo multiplex FM transmission standards were approved for commercial operation in the U.S. The National Stereophonic Radio Committee has been formed by the radio manufacturing industry in 1959 to test various proposals. The system developed by General Electric was recommended to and approved by the FCC in early 1961. Stereo transmission is downward compatible, meaning it allows monaural reception in radios not equipped with stereo reception.

Digital

AM and FM transmissions are analog signals, subject to an inherent background electrical noise (or ''hiss''), although FM suffers less than AM. Most consumers were introduced to digital sound and its superior signal-to-noise ratio with the success of the digital compact disc in the 1980s, and began to seek similar quality over the air.

By the late 1990s, most American and many international radio stations used the Internet to stream their signals, making worldwide reception possible. Audio streaming quality of sound was most often limited by the computer speakers used.

Based on research that began in the early 1980s, digital audio broadcasting (DAB) became operational in Europe and Canada, which in the 1990s agreed to use the ''Eureka 147'' technical standard, allowing transmission of digital-quality sound as well as information and data. Depending on the country, stations operated on Band III (around 221 megahertz) or L-band (1452 to 1492 megahertz) frequencies, well above current FM/VHF frequencies. L-band was allocated for digital radio at the World Administrative Radio Conference in 1992. The first commercial DAB receivers became available in 1998; and terrestrial DAB service was available by late 2001 to nearly 300 million people from more than 400 digital radio stations, with plans for satellite delivery in the future.

Declining to adopt Eureka 147, U.S. manufacturing companies formed the National Radio Systems Committee (NRSC) to determine which of a half-dozen schemes to recommend for approval by the FCC. To ease the eventual transition to digital radio from existing analog stations, the industry and FCC sought an in-band, on-channel (IBOC) scheme whereby the new service would operate during a transition period side-by-side the analog stations that it would

eventually replace. However, agreement on a workable technical standard was continually delayed. In the meantime, digital audio radio service (DARS), transmitted from domestic satellites, was authorized by the FCC in the late 1990s. XM Satellite Radio began to offer its subscription-based 100-channel service in late 2001 while competitor Sirius planned to begin advertising-free transmissions by early 2002. Both required the purchase of special digital receivers.

See also **Radio, Early Transmissions; Radio Receivers, Valve and Transistor Circuits; Radio Transmitters, Continuous Wave**

CHRISTOPHER H. STERLING

Further Reading

NAB Engineering Handbook, 9th edn. National Association of Broadcasters, Washington D.C., 1999.
Inglis, A.F. *Behind the Tube: A History of Broadcasting Technology and Business*. Focal Press, Stoneham, MA, 1990.
Lessing, L. *Man of High Fidelity: Edwin Howard Armstrong*. Lippincott, Philadelphia, 1956.
McNicol, D. *Radio's Conquest of Space*. Murray Hill Books, New York, 1946. Reprinted by Arno Press, 1974.
Pawley, E. *BBC Engineering 1922–1972*. BBC, London, 1972.
Sterling, C.H. and Kittross, J.M. *Stay Tuned: A History of American Broadcasting*, 3rd edn. Lawrence Erlbaum Associates, Mahwah, NJ, 2002.

Useful Websites

Digital Radio: http://infoweb.magi.com/~moted/dr/
World Forum for Digital Audio Broadcasting: http://www.worlddab.org/

Radio, Early Transmissions

In the transition from experiments to regular broadcasting, important pioneering efforts in radio transmission took place before the beginning of World War II. Records conflict, and there is some controversy over the primacy of some "firsts." Some pioneers have likely been forgotten for lack of the promotion available to others. Most of the experimental precursors transmitted telegraphy code; only slowly did the ability to send speech and music through the air become possible.

Wireless Telegraphy

Any consideration of pioneering transmissions must begin with German physicist Heinrich Hertz who, in a series of experiments in late 1887 and early 1888, was the first to demonstrate that prior theorizing about wireless was indeed correct. He succeeded in transmitting telegraph code across a room (using what soon became known as "Hertzian waves") without wire connections. Though he was not interested in the commercial potential of his findings, their publication triggered research by many others.

Guglielmo Marconi learned of his work while reading Hertz's obituary in early 1894. Within a year, using Hertz's spark-gap technology with the important addition of an antenna wire, Marconi was transmitting telegraphy signals for several hundred meters on his father's estate near Bologna, Italy. By 1896 he was in England, demonstrating wireless transmissions over several kilometers to government officials. Transmissions across the English Channel followed on March 27, 1899. On December 12, 1901, Marconi and several assistants (see Figure 5) succeeded in transmitting the Morse code signal for the letter S (three dots) across the North Atlantic from Poldhu in Cornwall to Signal Hill in St. Johns, Newfoundland. This was the first public demonstration of the feasibility of long-range radio communications, though regular commercial transatlantic wireless telegraph services did not commence until October 17, 1907.

That wireless would play an important role at sea was acknowledged early on. When the East Goodwin lightship in the English Channel was struck during heavy fog in 1898, a wireless transmission from the vessel helped to save lives—the first of many such radio feats. Regular radio transmissions to and from naval vessels began with Royal Navy operations in 1900, the same year Marconi first equipped several German liners with wireless transmitters. After several years of experimentation, regular Marconi commercial wireless telegraph transmission to and from merchant shipping began in 1904. Other maritime rescues using wireless became common, including the January 23 1909 saving of more than 1500 people from the *S.S. Republic*. However, the April 14–15, 1912, *Titanic* disaster and the role of wireless in saving some 700 survivors captured the public imagination as to the benefits and potential of wireless transmissions and hastened developments. It also ensured adoption of regulations that required passenger vessels to have a radio operator on duty at all times.

In October 1899 the British Army first used wireless transmission during the Boer War in South Africa. Both the Russians and the Japanese made extensive use of wireless transmis-

Figure 5. Marconi and his assistants launching the kite-supported aerial at Signal Hill, St. John's, Newfoundland, December 1901. [*Courtesy of the Marconi Corporation.*]

sions, at sea and on land, during the brief but fierce Russo–Japanese War of 1904–1905. By World War I, wireless transmission was widely used by all combatants, though wireless was still supplementing wired telegraph and telephone connections.

Broadcasting: 1906 to 1941

By connecting a telephone to a high-frequency alternator of his design, Reginald Fessenden became the first to transmit speech and music on Christmas Day and again on New Year's Eve of 1906, from Brant Rock, Massachusetts south of Boston. His success (which followed wireless telephony experiments dating back to 1902) with an "audience" of shipboard radio operators was widely publicized at the time and was among the first indications that wireless might move beyond telegraph signals. In the next few years, there were many other one-time broadcast demonstrations. Audion inventor Lee de Forest transmitted speech and music in 1908 from the Eiffel Tower in Paris and the voice of Enrico Caruso from New York's Metropolitan Opera on January 13, 1910. The first transmissions to and from an airplane in flight took place in both Britain and the U.S. in 1910; the first airliner equipped with radio flew the London to Paris route a decade later. On the West Coast of the U.S., Charles D. Herrold began occasional, and soon regularly scheduled, voice and music broadcasts from a transmitter at his radio school in San Jose, California. The outbreak of World War I in 1914 put an end to all of this experimentation.

Radio broadcasting resumed only after hostilities concluded in late 1918. During 1919 and 1920, initial and occasional broadcast transmissions of speech and music emanated from stations in The Netherlands, Canada, Britain, and the U.S., most of them initiated by amateur radio operators. Marconi began daily broadcasts from Chelmsford, England, on February 23, 1920. While the election night broadcast of November 2, 1920, from station KDKA in Pittsburgh, Pennsylvania, is often credited with being the first regular broadcast transmission, several other stations can make the same claim and cite earlier dates. The first coast-to-coast broadcast in the U.S. on October 24, 1924, combined telephone wires to transmit a presidential speech from New York to 22 stations across the country, previewing the rise of regular commercial network transmissions just two years later.

The first short-wave radio transmissions were undertaken by Marconi researchers in the early 1920s. Amateur operators sent the first such signal across the Atlantic on December 11, 1921. Regular two-way amateur traffic began on December 8, 1923, and has continued since. Westinghouse used short-wave transmissions to interconnect U.S. broadcasting stations in 1923–1924. By the end of the decade, the first regular cross-border, or international, short-wave propaganda broadcasts were being transmitted by the Soviet Union. Soon other nations began their own such services, including the BBC's Empire Service on December 19, 1932. On January 7, 1927 regular commercial transatlantic radio-telephone trans-

missions began between London and New York using AT&T and British Post Office long-wave circuits. These were supplemented with short wave in early 1927, and full two-way short-wave links opened June 1, 1929, soon replacing the long-wave transmissions. These were the first regular voice telecommunication links across the Atlantic to operate commercially and, updated from time to time, were the only such link until the first submarine telephone cable was opened for service in 1956.

The first public transmission of frequency modulation (FM) radio (prior developments were in AM, or amplitude modulation, radio) came on November 6, 1935, when U.S. inventor Edwin Howard Armstrong demonstrated his new system to a meeting of engineers at the Institute for Radio Engineers in New York. While commercial FM broadcasts began in the U.S. on January 1, 1941 (and on their present VHF spectrum band in 1945), FM stations in Europe appeared only slowly after the war, operating first in Germany. FM took decades to become a commercial success, achieving substantial audiences only in the 1970s in both Europe and the U.S.

See also **Radio: AM, FM, Analog, Digital; Radio Transmitters, Continuous Wave; Radio Receivers, Early; Radio Transmitters, Early; Radio Receivers, Valve and Transistor Circuits**

CHRISTOPHER H. STERLING

Further Reading

Aitken, H.G.J. *The Continuous Wave: Technology and American Radio, 1900–1932*. Princeton University Press, Princeton, 1985.

Dalton, W.M. *The Story of Radio, Part I: How Radio Began*. Adam Hilger, Bristol, 1975.

Douglas, S.J. *Inventing American Broadcasting, 1899–1922*. Johns Hopkins University Press, Baltimore, 1987.

Fleming, J.A. *The Principles of Electric Wave Telegraphy and Telephony*, 2nd edn. Longmans, London, 1910.

Hong, S. *Wireless: From Marconi's Black-Box to the Audion*. MIT Press, Cambridge, MA, 2001.

Maclaurin, W. R. *Invention and Innovation in the Radio Industry*. Macmillan, New York, 1949. Reprinted by Arno Press, 1971.

McNicol, D. *Radio's Conquest of Space*. Murray Hill Books, New York, 1946. Reprinted by Arno Press, 1974.

Useful Websites

A site sponsored by Marconi Corporation that explores the life, science and achievements of Guglielmo Marconi: www.marconicalling.com/

Radio Receivers, Coherers and Magnetic Methods

From 1900 to about 1914, when Morse code was the normal language of radio communication, a variety of methods for the detection of electromagnetic radiation were employed. Two of the more successful utilized discoveries made during the nineteenth century, namely (1) the property possessed by some types of loose contacts to cohere when activated by radio-frequency currents and go from a non-conducting to a conducting state and (2) the phenomenon of hysteresis in the magnetization and demagnetization of magnetic materials. Detectors based on these properties were in common use until their gradual replacement by crystal diodes and by vacuum tubes (or valves), which were becoming common components of electromagnetic wave detector circuits just before World War I. Coherers and magnet materials were not in themselves detectors, but they acted in conjunction with other circuit elements to render radio waves sensible to human beings by producing visible or audible signals. While coherers could be made to operate galvanometers, printers, and earphones, the output of magnetic detectors was sufficient only to operate earphones.

In 1890 the French experimenter Edouard Branley was the first to note that metallic powders, iron filings for example, exhibited the property of coherence. Subsequently, several workers investigated the phenomenon and produced their own devices, examples of which are shown in Figure 6. Some of the more significant contributions to the development of coherers were made by Oliver Lodge in Britain and by Alexander Popov in Russia, and both men devised their own circuits to be used in conjunction with their detectors. There was no standard form of coherer, nor did there appear to be particular suppliers of the devices. Many types were developed, in fact almost as many as there were experimenters, and they ranged in structure from solid materials in contact to powders. They were often precisely engineered, although it is not clear that precision was really required. Experimenters had to take into account the fact that once the coherer was put into the conducting state it did not return to its starting condition of nonconduction. In many devices the original condition could be obtained only by mechanical interference, shaking or striking the coherer for example. Some elaborate arrangements were used to achieve correct and reliable operation and to overcome the disadvantage of slow operating speeds that were contingent upon a mechani-

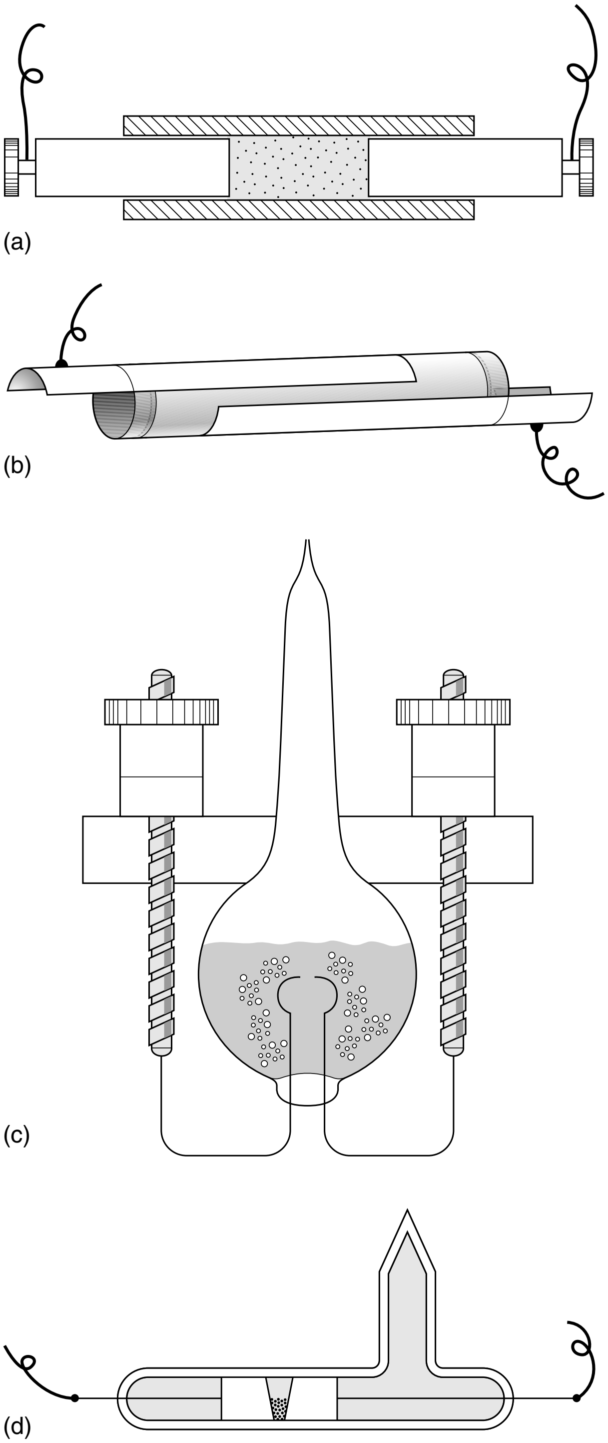


Figure 6. Types of filings coherers: (a) Branley, 1890; (b) Popov, 1895; (c) Lodge, 1895; and (d) Marconi, 1895. [*Source: Phillips, V.J. Early Radio Wave Detectors, IEE, London, 1980, p. 31.*]

cally operated device. If wireless telegraphy were to be a serious competitor to the existing terrestrial and submarine telegraph cables, then high signaling speeds would be essential. Terrestrial telegraphy could operate at hundreds of words per minute, and even very long submarine cables could carry messages at scores of words per minute. Mechanically operated coherers found it difficult to achieve speeds greater than 10 to 20 words per minute, which did not present serious competition to existing means of telegraphy.

There were some materials that displayed the properties of coherence and returned without external action to the nonconducting condition. Substances used in such self-restoring or autocoherers were most commonly steel, carbon, and mercury in various combinations (Figure 7). Other such devices were constructed by instrument makers such as E. Ducretet of Paris. In spite of great interest in coherers and their common use as radio detectors, no satisfactory theory of their operation has been given.

There were several methods of using coherers, and a typical circuit is shown in Figure 8. The battery served two purposes: to provide current through the galvanometer and earphone and to keep the coherer in a high-resistance, or sensitive, condition. In this state very little current flowed around the circuit abcd. However, when a radiofrequency signal was present in the aerial circuit, the resistance of the coherer dropped significantly, resulting in a sudden increase in the current, which caused the galvanometer to deflect or to be heard as a click in the earphone.

Magnetic detectors, a second type of early detectors, fell into two broad groups: (1) electrodynamic detectors, in which the presence of a radio signal was shown by the movement of a needle or a mirror, or (2) magnetic hysteresis detectors, in which the oscillatory field changed the state of magnetization of a piece of ferromagnetic material due to an aerial current. A common form of the magnetic hysteresis detector was that developed by Guglielmo Marconi (Figure 9). A loop of iron wire driven by a clockwork motor passed under two magnets and was cycled around a hysteresis loop (see Figure 10). Any oscillating decreasing currents in the aerial coil would tend to move the iron to a

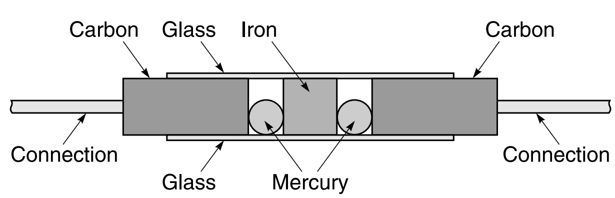


Figure 7. Auto-coherer.
[*Source: Adapted from Sewall, C. H. Wireless Telegraphy, Crosby Lockwood, London, 1904, p. 161.*]

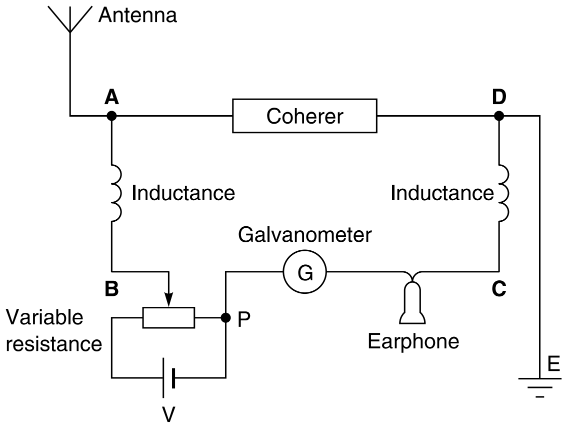


Figure 8. Basic coherer circuit.
[*Source: Adapted from Phillips, V. J. Early Radio Wave Detectors. IEE, London, 1980, p. 20.*]

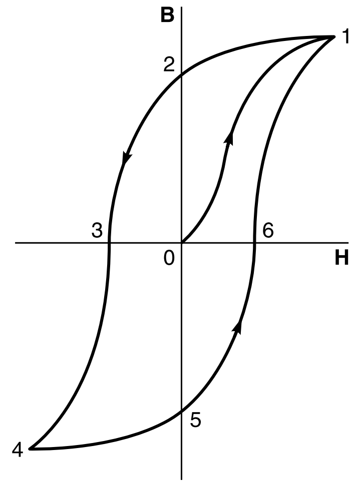


Figure 10. Hysteresis loop.

magnetization state corresponding to the path 1 to 0, and the change would be heard as a signal in the earphone via the shorter secondary coil surrounding the wire.

Before World War I, radio communication had largely been restricted to ship-to-shore communication and to military uses, and coherers and magnetic detectors were adequate for these applications. Although widely used for the reception of Morse signals, these devices were unsuitable for the detection of continuous waves. In the first decade of the twentieth century, continuous waves had been generated, typically by high-speed alternators. Although speech had been broadcast, these experiments had been desultory. During World War I, much attention was devoted to developing wireless telephony, and vacuum tubes were found to be the most effective way of generating continuous radio waves that could carry speech. After the war, particularly when the market for radio widened in the 1920s to become a fully commercial consumer-oriented industry, coherers and magnetic detectors vanished to be replaced by vacuum tubes and the crystal detector.

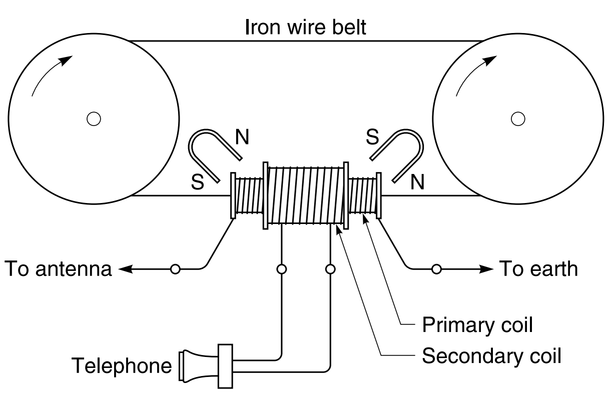


Figure 9. Schematic drawing of Marconi's hysteresis detector.
[*Source: Adapted from Dalton, W. M., The Story of Radio, Part 1: How Radio Began. Adam Hilger, Bristol, 1975, p. 95.*]

See also **Radio Receivers, Crystal Detectors and Receivers; Radio Receivers, Valve and Transistor Circuits; Radio Transmitters, Early; Rectifiers; Vacuum Tubes/Valves**

COLIN HEMPSTEAD

Further Reading

Beauchamp, K. *History of Telegraphy*. Institution of Electrical Engineers, London, 2001.

Burns, R.W. Discovery and invention: lodge and the birth of radio communications. *IEE Review*, 40, 131–33, 1994.

Dalton W.M. *The Story of Radio, Part I: How Radio Began*. Adam Hilger, Bristol, 1975.

Fleming J.A. *Fifty Years of Electricity: the Memories of an Electrical Engineer*. Wireless Press, London, 1921.

Phillips V.J. *Early Radio Wave Detectors*. Peter Peregrinus, Stevenage, in association with the Science Museum, 1980.

Phillips, V.J. The "Italian Navy Coherer" affair: a turn of the century scandal. *Proc. IEEE*, 86b, 248, 1998.

Pickworth, G. Coherer-based radio. *Electronics World and Wireless World*, 100, 563–567, 1994.

Radio Receivers, Crystal Detectors and Receivers

Crystal detectors based on the phenomenon of rectification were commonly employed in commercial radio receivers in the years following World War I. They were very effective in the detection of

modulated radio waves that carried speech and musical signals; and, because they could be employed directly to operate earphones, they were particularly useful in locations where a supply of electricity was not easily available. A circuit consisting of a crystal rectifier, a variable capacitor in parallel with an inductance, and an earphone, was sufficient to detect signals from commercial transmitters (Figure 11). However, even with a modicum of tuning they were lacking in selectivity; that is, the ability to separate transmissions from stations that were close in frequency (Figure 12).

In the 1870s several researchers noted the phenomenon of "unilateral conductivity," now termed "rectification." In 1874 Ferdinand Braun discovered that the electrical resistance of metal sulfides was direction dependent, and Arthur Schuster found a similar effect in circuits of copper and brass (where copper wires, for example, were connected with brass terminals). Desultory work continued through the last years of the nineteenth century, and more examples of rectification were found. Between 1900 and 1910 several systematic studies were made of crystal rectifiers with Henry Dunwoody of the U.S. Army patenting a device using carborundum. Between 1908 and 1910 Wichi Torikata and E. Yokoyama in Japan, and George W. Pierce and L.W. Austin in the U.S. added a large number of elemental and compound semiconductors that could be used as crystal detectors. It is clear that all crystal rectifiers utilized surface contacts either between two semiconductors or between fine metal points and semiconductors. The structure of a metal point and a crystal became known as a "cat's whisker." Simple crystal rectifiers were used commercially until well into the 1950s, although the cat's whisker had been replaced by then with encapsulated, fixed crystals, which were developed during World War II. The typical form of a crystal detector was a relatively large crystal against which could be placed either a thin wire (the cat's whisker) or a smaller crystal (see Figure 13).

Crystal sets were widely available in the early days of radio broadcasting when the technology became available to the home market. Nevertheless, despite the unreliability of devices that depended on the delicate physical contact between different materials, it was these that brought radio to a relatively large audience in the industrialized world. A large number of manufacturers produced crystal sets, some elaborate in their construction. Prices varied depending on the quality of the cases, for there was very little difference in the performance of the receivers. The quality of the cases was cosmetic rather than electronic in purpose. Nevertheless, many amateurs constructed their own crystal sets from the 1920s. Although crystal rectifiers were capable of possessing excellent characteristics, reliable results often depended on the user to apply considerable skill and patience in finding the best point of contact between the cat's whisker and crystal, or between two crystals. The sound quality was poor, and it was not easy for more than one listener to use a receiver, although acoustic amplifiers were sometimes employed. One homely method of amplification was a set of earphones placed in a bowl so that faint sounds could be heard by several listeners.

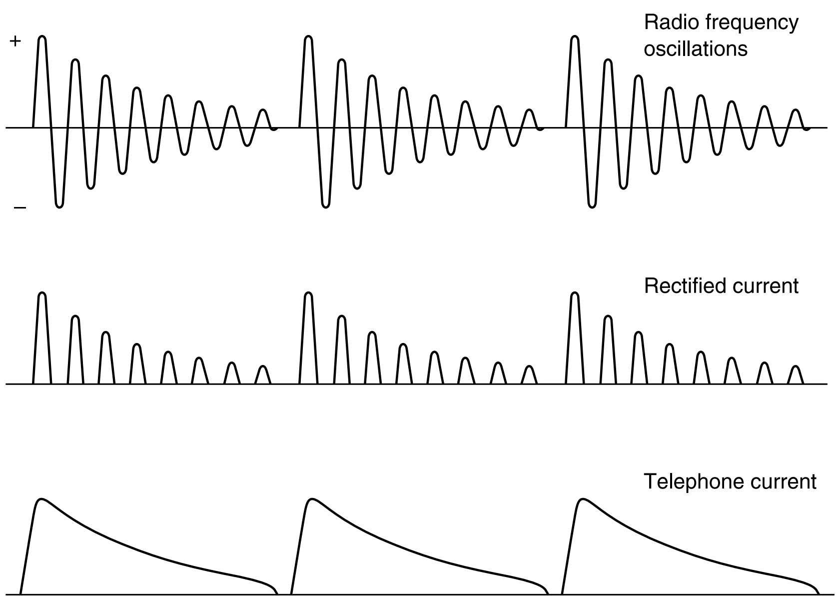


Figure 11. Rectification in action. [*Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy. Wireless Press, New York, 1917, p. 132.*]

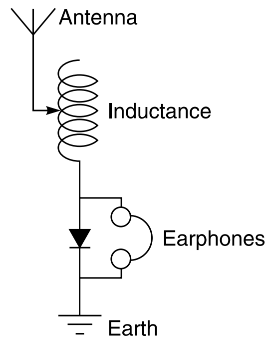


Figure 12. Simple crystal receiver.
[*Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy. Wireless Press, New York, 1917, p. 132.*]

By 1917 a wide range of detectors was available, most of which were of limited applicability and many were unsuitable for domestic purposes. Table 1 lists many of the types available and indicates their general utility.

Table 1 Types of crystal detectors.

	Detector Type
No local power	Galena; Zincite-Bornite; Carborundum; Fleming Diode
Local power	Carborundum, Zincite-Bornite; Fleming Diode; Triode; Silicon
Rectification	Galena; Silicon; Carborundum; Cerusite; Zincite-Bornite; Fleming Diode; Triode; Electrolytic
Detector: damped oscillations	Galena; Silicon; Zincite-Bornite; Carborundum; Fleming Diode; Triode; Marconi Magnetic
Detector: undamped oscillations	Tikker; Tone Wheel; Heterodyne Receiver; Vacuum Tube Oscillator

Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy. Wireless Press, New York, p. 141.

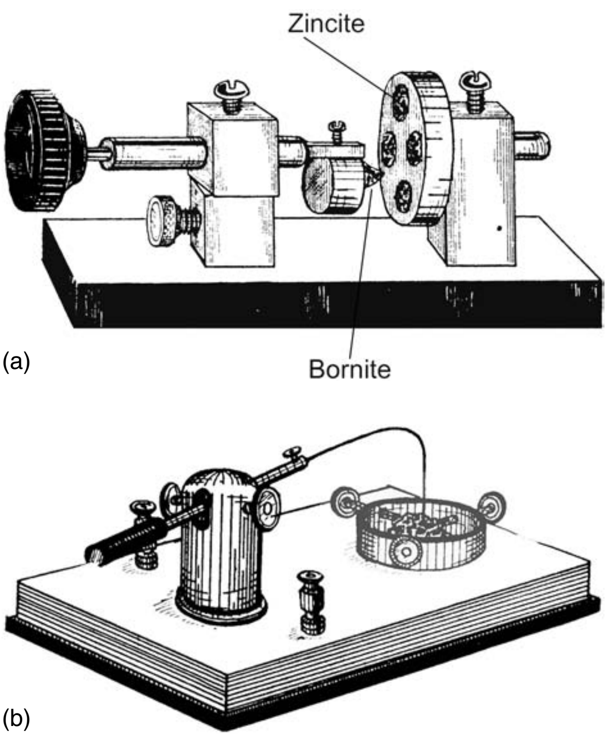


Figure 13. Two forms of crystal detectors. (a) a pointed crystal of Bornite touches one of four flat crystal of Zincite. (b) a fine wire touches flat crystals of silicon or galena. Each detector allows the most sensitive region to be found; on the left, different flat crystal can be brought in contact with the Bornite, and the pressures between the two crystals can be changed; on the right, a cats whisker contact can be adjusted while in contact with crystals of silicon or galena. [*Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy. Wireless Press, New York, 1917, p. 140.*]

The simple circuit was modified because while it was an effective detector, it was not particularly sensitive or discriminating. Therefore, a variety of inductively and conductively coupled crystal receivers were developed. Basic examples of these are shown in Figure 14. Some crystal detectors required the application of biasing voltages—

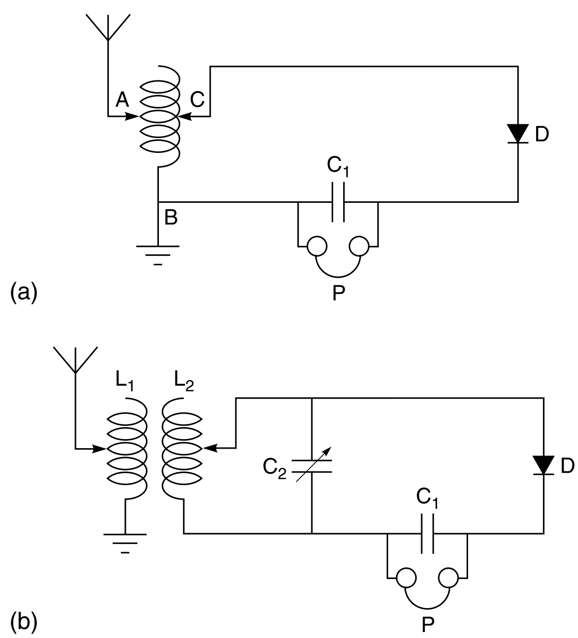


Figure 14. Coupling circuits: capacitance and inductance. [*Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy. Wireless Press, New York, 1917, p. 133.*]

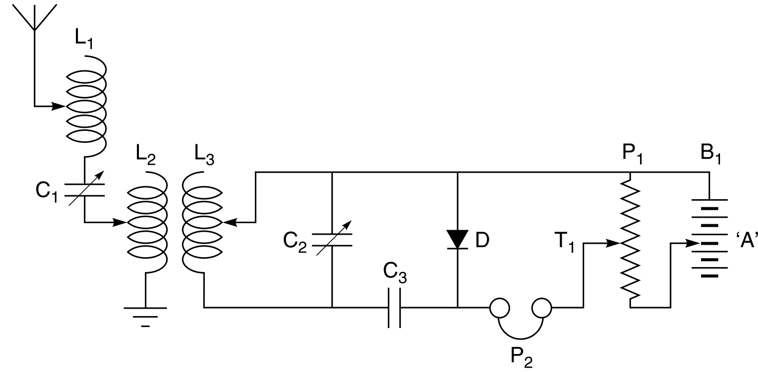


Figure 15. Biasing circuit used with carborundum. [*Source: Adapted from Bucher, E. E. Practical Wireless Telegraphy, New York, Wireless Press, 1917, p. 134.*]

carborundum in particular operated well with a current flowing through the junction—and a number of circuits were developed. Figure 15 illustrates one such circuit.

Crystal receiving sets were an important element in the development of public broadcasting. They were simple in construction, robust (although tricky to use), and above all, relatively cheap to buy. However, in spite of much work, effective theories of the phenomena of rectification were elusive. It was made evident very early in the history of crystal rectifiers that surface properties determined their actions, or at least the rectifying effects lay close to the edges of contacts. However, only when the modern theory of semiconductors was put forward in the 1930s did it become clear that chemical and crystalline purity were essential if reliable solid-state devices were to be made. A group of new technologies arose—zone-refining, solid-state diffusion, and so on—which allowed crystal rectifiers to be fully understood and to be subsumed into the modern industry of semiconductors.

See also **Radio Receivers, Coherers, and Magnetic Methods; Radio Transmitters, Early; Radio Receivers, Valve and Transistor Circuits; Rectifiers; Transistors; Valves/Vacuum Tubes**

COLIN HEMPSTEAD

Further Reading

Beauchamp, K. *History of Telegraphy*. Institution of Electrical Engineers, London, 2001.

Blake, G.G. *History of Radio Telegraphy and Telephony*. Radio Press, London, 1926.

Burns, R.W. Discovery and invention: lodge and the birth of radio communications. *IEE Rev*., 40, 131–133, 1994.

Bucher, E.E. *Practical Wireless Telegraphy*. Wireless Press, New York, 1917.

Dalton, W.M. *The Story of Radio*, *Part 1*, *How Radio Began*. Adam Hilger, Bristol, 1975.

Erskine-Murray, J. *A Handbook of Wireless Telegraphy*. Crosby, Lockwood, London, 1913.

Fleming J.A. *Fifty Years of Electricity: the Memories of an Electrical Engineer*. Wireless Press, London, 1921.

Phillips V.J. *Early Radio Wave Detectors*. Peter Peregrinus, Stevenage, in association with the Science Museum, 1980.

Radio Receivers, Early

While early magnetic detectors such as coherers provided an effective means of receiving wireless telegraph (code) signals, the continuous waves used to transmit wireless telephony (voice, music) required both improved means of reception and effective signal amplification. The crucial and related inventions that allowed this appeared within two years of each other early in the twentieth century, and led to considerable legal wrangling over patent control.

Edison Effect (1883)

While experimenting with his new incandescent electric light, prolific American inventor Thomas A. Edison noted that the glowing (because it was heated) lamp carbon filament somehow created conditions that slowly blackened the insides of light bulbs, eventually rendering them useless. Preoccupied with perfecting his breakthrough invention, he sought to understand and limit the causes (he presumed carbon particles had been thrown off), filed a patent on a related device in late 1883 (now considered to be the first electronics patent), but did not pursue practical applications of his finding.

What became known as the "Edison effect" soon fascinated researchers on both sides of the Atlantic, one of whom was John Ambrose Fleming, a London-based electrician for the British branch of Edison's company. Edison demonstrated the effect to Fleming while the latter visited the U.S. Fleming and others tried to determine what caused the effect and whether it had any potential use, and published

brief reports on their limited findings. All noted that a narrow strip of glass was not blackened, where one leg of the light filament shielded the other.

By the late 1890s, building on a growing understanding of both conduction through gasses and emissions from hot elements within vacuums, researchers finally understood that streams of electrons, not carbon particles, were responsible for the Edison effect.

Fleming Valve (1904)

Within a few years, devices to improve wireless technology were often the focus of experimental electrical work in many countries. Ambrose Fleming was one of many researchers who began to seek whether newly discovered electrons could be effectively harnessed to detect wireless signals. He had a particular reason for seeking an improved detector that could display its results on an electrical meter—he was partially deaf and often missed weak signals.

Fleming and others determined that electrons were negatively charged, and thus a positively charged plate could be used to attract them. Put another way, electrical current would "flow" only when a plate was positively charged. Inserting such a plate into a small glass vacuum tube along with a filament conductor created what looked very much like an incandescent light, but was in reality a one-way gate or "valve" which could change (rectify) alternating current (such as wireless or radio waves) to pulsating direct current. Using such a two-element (filament and plate) valve or "diode," an operator could detect the presence of wireless telegraphy signals.

Fleming (by then a scientific advisor for the Marconi Company in London) was granted British, American and German patents for his "oscillating valve" device in 1904–1905. Though his valve worked as well as the earlier magnetic detectors, and was somewhat more stable in operation, it demonstrated no greater sensitivity to weak signals. It was simply an alternative approach to wireless signal detecting. The Marconi Company made wide use of the device in many of its radio stations up to World War I. Fleming was knighted in 1929. Over time, his device (indeed all that followed it) became better known as "thermionic" valves as they involved the use of incandescent (heated) electrons.

De Forest Audion (1906)

In America, experimenter Lee de Forest (1873–1961) was also seeking improved means of wireless signal detection, especially those that might avoid further patent infringement problems that already plagued his operations. He had laboriously developed a series of detectors (one even used small gas flames) and soon turned to variations on the Fleming valve as a promising approach. However, he lacked a scientific understanding of its principals, as over the space of several months he sought to improve its operation by experimenting with added elements, soon focusing on placing an additional electrode inside the tube. This took the form of a second plate or "wing," and was patented in 1906. The resulting device was little better than Fleming's valve. Indeed, Fleming became bitter over de Forest's ignoring of his prior work (de Forest claimed not to know of the Fleming valve patent before 1906) as well as his own lack of any royalties on the diode (as the Marconi Company owned the patent rights).

De Forest's added tube element soon changed, taking the form of a tiny inserted wire in the form of a grid, and the true three-element tube or "triode" was born, dubbed the "Audion" by one of his assistants, and patented in 1907 (confusingly, the term Audion was applied by de Forest to many diode and triode devices). Even this latest version was initially perceived as just another detector among many available options.

The new Audion tube would, however, turn out to be a far more capable device—indeed it can be argued that it lay the foundation for electronics development until the era of solid-state devices; but realization of the full potential of the Audion took time, and much of the effort to make that so came from others. The improved tube's ability to amplify weak signals came to the attention of the American Telephone & Telegraph Company (AT&T), which for years had sought an amplifier for long-distance wired telephone circuits. AT&T purchased partial rights to the Audion in 1909 and using a modified Audion device, opened transcontinental telephone service in 1914.

Meanwhile the man who would become de Forest's primary radio rival, Edwin Howard Armstrong (1890–1954), had discovered the "feedback" potential of the Audion as a student at Columbia University. When an Audion's output was fed back into its own input, the circuit could become a generator of electrical signals—effectively a radio transmitter. While virtually all engineers credited Armstrong's genius, the U.S. Supreme Court terminated two decades of patent battles in 1934 when it held that de Forest deserved credit.

See also **Radio Receivers, Coherers and Magnetic Methods; Radio Receivers, Crystal Detectors and Receivers; Radio Receivers, Valve and Transistor Circuits; Radio Transmitters, Continuous Wave; Vacuum Tubes/Valves**

CHRISTOPHER H. STERLING

Further Reading

Aitken, H.G.J. De Forest and the Audion, in *The Continuous Wave: Technology and American Radio, 1900–1932*. Princeton University Press, Princeton, 1985.

Blake, G.G. The thermionic valve and some of the fundamental valve circuits, in *History of Radio Telegraphy and Telephony*. Chapman & Hall, London, 1928. Reprinted by Arno Press, 1974.

Chipman, R.A. De Forest and the triode detector. *Scientific American*, 212, 92–100, 1965.

Fleming, J.A. *The Thermionic Valve and its Developments in Radiotelegraphy and Telephony*. *Wireless Press*. London, 1919.

MacGregor-Morris, J.T. *The Inventor of the Valve: A Biography of Sir Ambrose Fleming*.Television Society, London, 1954.

McNicol, D. The magic bulb, in *Radio's Conquest of Space*. Murray Hill Books, New York, 1946. Reprinted by Arno Press, 1974.

Shiers, G. The first electron tube. *Scientific American*, 220, 104–112, 1969.

Thermionic Valves 1904–1954: The First Fifty Years. Institution of Electrical Engineers, London, 1954.

Tyne, G.F.J. *Saga of the Vacuum Tube*, 2nd edn. Sams, Indianapolis, 1977.

Radio Receivers, Valve and Transistor Circuits

The direction of radio receiver circuit development between the 1920s and the early twenty-first century has been closely linked to two interrelated factors. The first is the inseparability of the development of radio components and radio hardware.

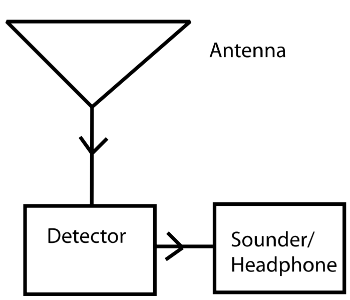


Figure 16. A simple radio receiver, consisting of an antenna, detector (such as a crystal diode, coherer, or other device), and a means of listening to the detected signals, often an electromechanical sounder or ink printer for wireless telegraphy.

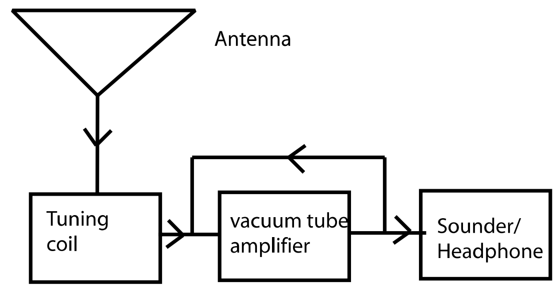


Figure 17. A regenerative receiver, utilizing an electronic vacuum tube amplifier with part of the output fed back to the input, thereby increasing the volume. By the time of the regenerative circuit, it was also necessary to include a tuner because of the large number of stations on air.

Improvements in components such as vacuum tubes and transistors have alternately stimulated or been stimulated by developments in circuit design. The second factor is the influence of social factors, sometimes seemingly unrelated to purely technical issues. Regulatory changes, for example, have determined how radio can be used and have been reflected in circuit designs. In some cases, such social factors have spelled success or failure for particular circuits, but more common is the phenomenon of national or international regulations extending the life of certain types of circuits by reigning in technological changes that would otherwise have made particular designs obsolete.

The earliest radio transmissions consisted simply of an electric spark generated through the use of a battery connected to a high-voltage electrical transformer (or coil), a capacitor, a means of switching current on and off, and miscellaneous other components such as resistors and interconnecting wires. The reception at increasingly greater distances of the resulting electromagnetic waves was of great concern to early radio engineers, stimulating many innovations in receiving circuits. Until the invention of the Fleming valve in 1904, most receiving circuits relied on electromechanical detectors of some kind, which had very limited sensitivity.

Basic spark transmitters in use around the turn of the twentieth century each broadcast a wide bandwidth signal, but these gave way to tunable transmitters because of the desire to conserve the electromagnetic spectrum. In fact, tunable transmitters were made necessary in order to comply with international treaties that attempted to reduce station interference. This change demanded circuits capable of restricting the frequency of the generated radio signal at the transmitter, as well as

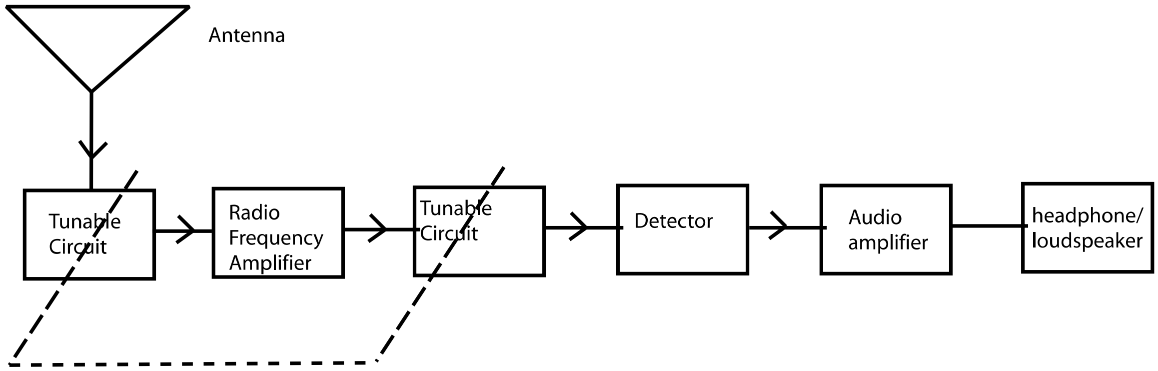


Figure 18. The neutrodyne circuit was the most successful variation of an earlier receiver design called the tuned radio-frequency (TRF) circuit. In a TRF receiver, the incoming signal passed through two or more stages of amplification before being detected. The tuning devices in each stage had to be mechanically linked together to operate in unison (shown by the dotted line). The neutrodyne overcame a major feedback problem by adding a "neutralizing" capacitor (not shown) in each radio frequency amplification stage.

circuits that would allow a receiver to select from a range of signals being broadcast at different frequencies. The latter task was almost universally accomplished by using a variable, air-gap capacitor in the receiving circuit so that the user could change the frequency at which the circuit received waves most efficiently.

Such receivers, using what was generically called a "tuned radio frequency" circuit, were common through the late 1920s. However, the advent of the amplifying vacuum tube in 1906–1907 radically changed the way radio circuits were designed: because the incoming signal could be electronically amplified, engineers had greater options. Receiver designers changed their emphasis from preserving the strength of the incoming signal to other issues, such as the ability to receive both strong and weak signals equally well. Key innovations of the vacuum tube era included the 1912 "regenerative" receiver of Edwin H. Armstrong of the U.S., which in effect tapped off a portion of the incoming signal and fed this back into the amplifying circuit, thereby creating a powerful feedback loop capable of great signal amplification (and hence capable of receiving weak, distant signals easily). Regenerative receivers were inexpensive, making them popular among amateur radio enthusiasts, but they were also unstable and could easily be overloaded, leading to unpleasant howling sounds. The major competitor of the regenerative receiver

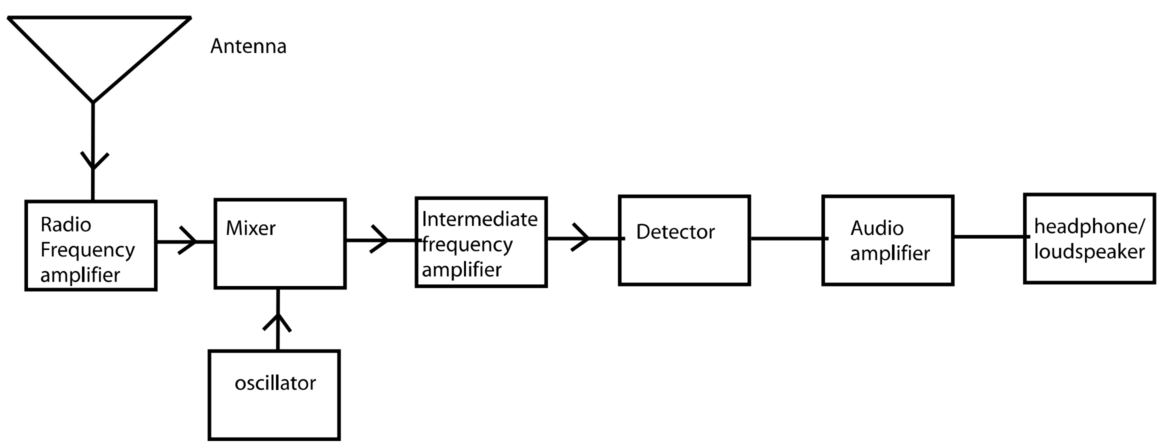


Figure 19. Superheterodyne receiver. The electronic oscillator produces a high-frequency signal, which is mixed or "heterodyned" with the incoming signal from the antenna. The product is kept within a narrow band of frequencies, allowing the following amplifier and detector stages to be optimized to a narrow bandwidth.

was the tuned radio frequency (TRF) receiver, which used several stages of electronic amplification to boost the incoming radio-frequency signal before feeding the signal to a detector. TRF receivers also tended to break into uncontrolled oscillation, but the ''neutrodyne'' circuit invented in 1922 by Harold Wheeler and Alan Hazeltine of the U.S.) used carefully chosen induction coils at key points in the circuit to counteract this tendency. The neutrodyne circuit, licensed by several radio manufacturers, did not have the problem of sudden loud howls caused by uncontrolled feed back loops as in the regenerative circuit.

Edwin Armstrong, however, had been busy at work while engaged in the U.S. Army Signal Corps during World War I. In 1918, he invented the superheterodyne circuit, which took a new approach to the issue of receiver sensitivity. While radio signals span a wide range of frequencies, radio tubes and circuits amplify some frequencies better than others. The superheterodyne first changed the frequency of the incoming signal, whatever it was, to a single ''intermediate'' frequency. The remaining receiver circuits, optimized to process signals in a band near the intermediate frequency, thus always had maximal sensitivity and performance. The superheterodyne circuit was widely licensed or copied in the 1930s and remains a standard feature of nearly every radio device.

By the 1930s, radio broadcasting authorities worldwide had settled on amplitude modulation (AM) as the standard form of radio transmission. Most countries worked to ensure that all their citizens were within the range of the radio networks. As part of this effort, a regulatory structure emerged that had the effect of petrifying AM technology. Other forms of modulating a radio wave had been proposed, but they were rejected for technical or economic reasons, or simply because it was desirable to support AM. The now-familiar frequency modulation (FM) form of broadcasting was proposed in the 1920s, but it was not until the 1930s that vacuum tube technology was sufficiently advanced to permit FM sound quality to exceed that of AM. Furthermore, implementing FM required a considerable political struggle because it competed with AM and the new technology of television.

Once again, Armstrong was at the forefront of the movement to reintroduce FM as a broadcast technology, demonstrating an improved system in 1934. Many countries around the world introduced FM broadcasting after World War II. Regulatory agencies gave FM broadcasters such wide swaths of bandwidth compared to AM stations that all sorts of additional services could be interleaved with ordinary audio transmissions. For example, many FM stations in the 1950s simultaneously broadcast radio-facsimile or background music services, which were not detectable on consumer receivers. These were abandoned in order to use the extra bandwidth for FM stereophonic broadcasts, which were approved by regulators as early as 1961 in the U.S.

The use of transistors had less of an effect on radio circuit design than might be expected. AM superheterodyne receivers using a minimal complement of five tubes were nearly universal by the 1950s, and in some cases transistors were substituted for tubes in virtually identical circuits. However, transistors are smaller and much more energy efficient than tubes, and their use resulted in home and portable receivers that were smaller and cooler in operation. The pocket-sized transistor radios that helped bring semiconductor technology to the attention of the public also used minimally modified versions of vacuum tube circuits. A memorable feature of the early transistor receiver designs was the installation of the transistors into plug-in sockets, mimicking vacuum tube designs. In the 1950s, the extremely high reliability of transistors was not yet known, so it was assumed that transistor replacement would be common and that repair would be made easier through the use of these sockets.

FM receivers required many more vacuum tubes than AM receivers, and hence were quite expensive before the introduction of the transistor. The price reductions possible with the use of transistors helped spur sales of FM receivers (which were not selling well before that) and hence popularize FM broadcasting. Many of the circuit innovations of the transistor period were related to the low cost of transistors, which encouraged manufacturers to add features to existing designs. In the 1970s, for example, signal processing tasks such as ''active,'' amplified tone control circuits replaced the simpler passive filter networks used earlier. Such devices as the elaborate graphic equalizer tone controls introduced as early as the 1960s would have been prohibitively expensive for the consumer market if rendered in vacuum tube circuitry. Since the 1960s, the integrated circuit has extended this trend, reducing the number of components in receiver, amplifier, and tone control circuits. It is possible today to put all or most of the circuit elements necessary to make a functional radio receiver on a single integrated circuit requiring only an external battery, antenna, and earphone or loudspeaker.

See also **Radio Receivers, Crystal Detectors and Receivers; Radio Receivers, Early; Radio Transmitters, Early**

DAVID MORTON

Further Reading

Aitken, H.G.J. *The Continuous Wave: Technology and American Radio, 1900–1932*. Princeton University Press, Princeton, 1985.

Morton, D. *A History of Electronic Entertainment Since 1945*. IEEE History Center, New Brunswick, NJ, 1999.

Radio Club of America. *The Legacies of Edwin Howard Armstrong*. N.p.: Radio Club of America, Inc.

Radio Transmitters, Continuous Wave

In the 1900s, most contemporary wireless telegraphy transmitters were spark-gap generators. Because of their high radiation resistance, the waveforms produced by a spark gap died out, or damped, quickly. While damped waveforms were not a problem for telegraphy, they presented a serious obstacle for telephony because continuous voice signals needed to ride on continuous waves. Continuous wave transmitters had to be developed for wireless telephony.

The arc generator was an early step in the development of continuous wave transmitters. The arc phenomenon appeared when two conductors previously joined were separated; the current kept flowing between the separated conductors, and a flame-like arc was established. In the 1890s in Britain, Hertha and William Ayrton discovered that an electric arc had a negative resistance. Connecting an arc to an ordinary oscillation circuit, the arc's negative resistance compensated for the positive radiation resistance and thus reduced the damping effect. The electric arc could therefore be used to construct a continuous wave oscillator. William Duddle, Ayrton's student, implemented this idea with his "singing arc" circuit, in which an arc lamp was connected with a capacitor and an inductor in parallel (Figure 20). This circuit could generate continuous waves up to 15 kilohertz. In 1903, the Danish physicist Valdemar Poulsen transformed the Duddle arc into a megahertz oscillator by replacing air with hydrogen gas, applying a magnetic field, using a water-cooled positive copper electrode, and rotating the carbon electrode. In the 1910s, the Poulsen arc became an important technology for wireless telegraphy, but it was not a genuine continuous wave transmitter. British engineer John Ambrose Fleming demonstrated that an arc generator produced a series of short pulses instead of continuous waveforms.

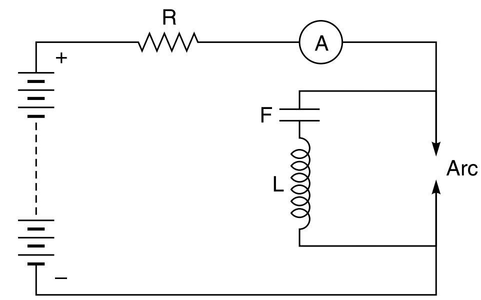


Figure 20. Duddle's singing arc circuit.
[*Source: William Duddle, J.Inst. Elec. Eng., 30, 232–283, 1900. Reproduced with the permission of the Institution of Electrical Engineers.*]

A promising candidate for a true continuous wave radio transmitter was the high-frequency alternator. The alternator, a technology for producing alternate electricity, used the principle of electromagnetic induction. By rotating an armature wound with coils in a fixed magnetic field, an electromotive force was induced in the coils. An alternator was typically used to generate high-power electricity below 150 hertz. In 1901, Reginald Fessenden proposed to generate radio-frequency waveforms with a high-speed alternator and to couple these waveforms directly to the antenna circuit. Unlike the spark-gap transmitter that produced damped waves, the alternator-based transmitter would provide fully undamped waves as long as its armature's motion was maintained. To design this kind of alternator, Fessenden contracted with the General Electric Company. The most critical technical challenge for its development was speed. A radio alternator required the armature to rotate at 50,000 cycles per second or more so that the signal frequency could exceed 50 kilohertz. At the time, no known mechanical armature could withstand the centrifugal force produced by such a high rotational velocity, but General Electric engineer Ernst Alexanderson was eventually able to overcome this problem. In his design, the rotary armature was replaced by two stationary armatures and a rotary steel disk, the circumference of which was embedded with iron teeth, or poles of an electromagnet. An electromotive force could be induced in the coils, and one could increase the frequency of the electromotive force by adding more poles on the disks' circumferences. Rapid disk rotation did not pose the mechanical problem of a rotating armature.

A schematic drawing of Alexanderson's alternator is shown in Figure 21. A magnetic flux, *M*, that is produced by a field coil, *S*, passes through

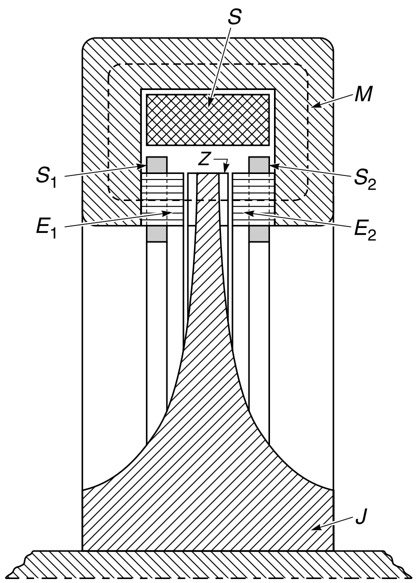


Figure 21. The Alexanderson alternator.
[*Source: Zenneck, J. and Seeling, A. E. Wireless Telegraphy. McGraw-Hill, New York, 1915, p. 214. Reproduced with the permission of McGraw-Hill.*]

the iron cores *E1* and *E2* of the armature coils *S1* and *S2*. The rotary iron disk, *J*, has teeth, *Z*, on its periphery that can extend to the space between the two iron cores *E1* and *E2*. When a tooth of the iron disk is in the gap between the iron cores, the overall reluctance of the magnetic circuit is small, so the magnetic flux, *M*, is large and the induced voltages on *S1* and *S2* are large. When there is no iron tooth between the cores, the air gap contributes a large reluctance to the magnetic circuit, so the magnetic flux, *M*, is small and the induced voltages on *S1* and *S2* are small. Therefore, as the disk rotates, the induced voltages vary periodically between a maximum value and a minimum, and the frequency of this variation is the number of teeth multiplied by the angular velocity of the disk.

Based on Alexanderson's design, General Electric successfully made a 2-kilowatt 100-kilohertz alternator in 1909 and delivered it to Fessenden. Following the 1909 machine, alternators with higher power and frequency were further developed. Although gradually replaced by tube transmitters, alternators had greater survivability in long-distance communications. The U.S. Navy's alternator in Hawaii remained operational until 1957. Another alternator in Massachusetts was used by the U.S. Air Force until 1961.

Another continuous-wave transmitter technology used vacuum tubes. Invented by Fleming and American Lee de Forest in the mid-1900s, vacuum tubes were first used in radio detection. Their potential in radio transmission was not explored until the early 1910s. By about 1912, de Forest, Fritz Lowenstein, and Edwin Armstrong in the U.S, H.J. Round in Britain, and Alexander Meissner in Germany had all conceived the idea of vacuum-tube oscillators. They discovered that when connecting the output of a tube amplifier with its input via a resistive–inductive–capacitive network an oscillatory electric current would appear at the circuit. The reason, de Forest argued, was that the amplifier and the feedback network made the entire circuit a regenerative system. Signals from the tube's output were directed to its input and were further enhanced by amplification. Because the feedback network directed only a narrow frequency band to the input, the self-regenerative wave was nearly sinusoidal. The explanation implied that the square of the resonating frequency of a tube oscillator was inversely proportional to the inductance and capacitance of the overall circuit. Therefore, by adjusting the inductors and capacitors in the feedback circuit, one could vary the resonating frequency of a vacuum-tube oscillator.

A diagram of a primitive vacuum-tube oscillator, proposed by Meissner in 1913, is shown in Figure 22. When a tiny resonance is produced at the circuit of *L1*, *L2*, and *C*, it is coupled to the vacuum tube's grid, *G*, through the inductor *L3*. The resonating signal is then amplified at the tube's plate, *P*. The amplified plate current is coupled to *L1* through the inductor, *L4*, and hence enhances the original resonance. Such a feedback cycle goes on and on until the resonating signal gains its maximum.

The vacuum-tube oscillator became the dominant radio transmitter after 1912. But when engineers began to explore waves beyond 10

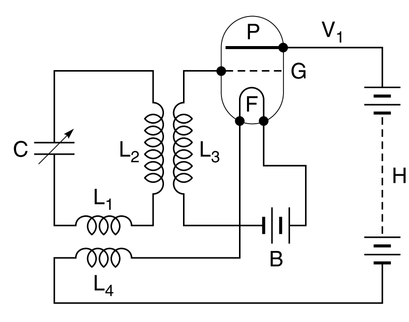


Figure 22. Meissner's oscillating circuit.
[*Source: Scott-Taggart, J. Thermionic Tubes in Radio Telegraphy and Telephony. Iliffe & Sons, London, 1924, p. 289. Reproduced with the permission of Iliffe & Sons.*]

megahertz during World War I, the de Forest-type vacuum tube oscillator was found to be inadequate because it was necessary to either shrink the tube's size or change the circuit structure in order to obtain high frequencies. In 1920, H. Barkhausen and K. Kurtz in Germany proposed raising the oscillator's frequency to 1500 megahertz by giving a strong positive bias to the tube's grid. The Barkhausen–Kurtz oscillator was the most popular ultrashortwave transmitter until the invention of microwave devices in the 1930s.

See also **Radio Receivers, Early; Radio Transmitters, Early; Radio Receivers, Valve and Transistor Circuits; Telephony, Long Distance; Valves/Vacuum Tubes**

CHEN-PANG YEANG

Further Reading

Aitken, H.G.J. *The Continuous Wave: Technology and American Radio, 1900–1932*. Princeton University Press, Princeton, 1985.

Brittain, J. *Alexanderson: Pioneer in American Electrical Engineering*. Johns Hopkins University Press, Baltimore, 1992.

Douglas, S.J. *Inventing American Broadcasting: 1899–1922*. Johns Hopkins University Press, Baltimore, 1987.

Fleming, J.A. *The Principles of Electric Wave Telegraphy and Telephony*. Longman, Green & Company, New York, 1916.

Hong, S. *Wireless: From Marconi's Black-Box to the Audion*. MIT Press, Cambridge, MA, 2001.

Scott-Taggart, J. *Thermionic Tubes in Radio Telegraphy and Telephony*. Iliffe, London, 1924.

Zenneck, J. and Seeling, A.E. *Wireless Telegraphy*. McGraw-Hill, New York, 1915.

Radio Transmitters, Early

According to James Clerk Maxwell's theory, the very existence of electromagnetic waves entailed the principle that matter can act on other matter at a distance without any intervening matter. Before Heinrich Hertz's experimental work with electromagnetic waves, no one was confident about this implication. Hertz proved it so by detecting the wave generated by a spark from an electrical discharge across the gap formed by two electrodes. Hertz was driven to this experiment by the challenge posed by Maxwell's theory, but as a theoretician himself, he had no interest in (and therefore could never have predicted) the application of his discovery—the spark transmitter and the early form of radio transmission.

Hertz was not the first to produce electromagnetic waves, but history regards him as the first to detect the waves he produced. In 1887 he created what he called "electric waves" by generating a sufficiently high voltage to create a spark. He detected the resulting electromagnetic wave by means of a loop of wire, which he called a "resonator," the ends of which terminated in a small gap. When the wave passed through the loop, an electric current was induced in the wire causing a spark to leap across the gap.

Hertz's discovery provoked much speculation about the use of electromagnetic waves for signaling, especially between land and ships in danger of running aground in fog. His apparatus, however, did not produce waves that traveled far enough or were stable enough for multiple simultaneous transmissions in the same locale. In fact, because Hertz analogized the propagation of electromagnetic waves to the propagation of light, he never would have overcome the limitations of his initial invention.

Early spark-gap transmitters consisted of a pair of electrodes and a combination of condenser and coil that were tuned to a certain frequency. The spark generated across the gap provided a short burst of electrical energy that shocked the tuned circuit into oscillation producing an electromagnetic wave that would eventually die out. Applying continuous shocks to the tuned circuit, however, would sustain the wave, much like a flywheel is kept in rotation by the continuous but intermittent force of a piston.

These transmitters produced waves that were inefficient, variable in frequency, and broadbanded. When a spark occurred across the gap, it caused some of the metal in the electrodes to vaporize. Being conductive, this vapor allowed an electrical arc to form across the gap, which lowered the efficiency of the transmitter and caused it to generate waves of multiple frequencies. Some means of dissipating the vaporized metal around the electrodes was needed—a process later called "quenching."

In 1889 Oliver Lodge demonstrated electrical resonance, the property that made possible the generation of a wave of one stable frequency. From that point the search was on to find the optimum combination of configuration and composition for the design of spark transmitter electrodes. Marconi's designs included iron electrodes shaped like mushrooms. A variety of other metals went into the electrodes: silver, aluminum, graphite, tungsten, and others. In an effort to quench the arc, some electrode designers tried forcing the spark through running liquids or encasing the electrodes in water-cooled jackets. Nikola Tesla and other researchers developed rotary gaps to

produce sparks with higher repetition rates and thus more stable waves. Such continuous waves could produce musical notes; in fact, some early wireless stations in Great Britain were known to have transmitted electromagnetic renditions of "God Save the King." As these designers produced purer sinusoidal waves, the prospects for transmitting voice seemed within their reach. No designer, however, could overcome the spark transmitter's inherent noisiness in order to make radio telephony practical.

Spark transmitter designs did eventually make radio telegraphy practical. In 1897, Marconi secured a patent in Great Britain for a wireless telegraphy system that could transmit messages up to 3 kilometers. Most of the components of his system were adaptations of existing devices—Righi's spark transmitter, a Branly coherer, and a Morse inker, among others. Marconi's contribution that broke the limitation on transmission distance was his decision to liken his wireless system to the existing wired telegraphy systems. Transmitters and receivers in a wired telegraphy system used an earth ground as one of their interconnections; Marconi grounded his spark-gap transmitter and its receiver, and thereby created the first practical wireless telegraphy system. Marconi went on to develop his spark transmitter system to produce a 5-centimeter spark from 100,000 volts that transmitted a series of Morse dots from Cornwall to Newfoundland—the first transatlantic radio transmission.

On the American side of the Atlantic, Lee de Forest and Reginald A. Fessenden were developing their own spark transmitter systems. De Forest's design consisted of tungsten electrodes with an adjustable gap; Fessenden's was a rotary spark-gap design. Both of these inventors attracted the attention of the U.S. Navy, whose patronage was a major force in the development of wireless telegraphy. At the same time they were also testing the Poulsen arc transmitter, which covered greater distances and was freer from noise.

Spark transmitters were pressed into service in the early years of the twentieth century in a variety of settings. The armed forces of many nations used them, and spark transmitters first played an important role in war in the conflict between Japan and Russia in 1904. Cruise and merchant ships at this time also used spark transmitters; the *Titanic* broadcast being one of the first distress messages in the form of SOS in 1912.

Ultimately, however, de Forest's invention of the triode vacuum tube, or thermionic valve, in 1912 made possible the efficient generation of pure sinusoidal electromagnetic waves and rendered the spark transmitter with its noisy broadbanded output obsolete. In its brief lifetime the spark transmitter was essential for jump-starting the theory and practice of radio communication. Today, however, it is an historical curiosity, having been banned from the ether in 1927 by U.S law because of concerns about fire safety.

See also **Radio, Early Transmissions; Radio Receivers, Coherers and Magnetic Methods; Radio Receivers, Crystal Detectors and Receivers; Radio Receivers, Early**

LAWRENCE SOUDER

Further Reading

Archer, G.L. *History of Radio*. New York, 1938.

Beauchamp, K. *History of Telegraphy*. London, 2001.

Blake, G.G. *History of Radio Telegraphy and Telephony*. London, 1926.

Hertz, H. *Electric Waves*. London, 1900.

Hong S. *Wireless: From Marconi's Black-Box to the Audion*. Cambridge, 2001.

Howe, G.W.O. *Handbook of Wireless Telegraphy and Telephony*. London, 1921.

Useful Websites

Belrose, J.S. The Sounds of a Spark Transmitter, Telegraphy and Telephony: http://www.hammondmuseumofradio.org/spark.html

Radioactive Dating

There are natural radioactive isotopes that have half-lives comparable to the age of the earth, the most familiar being uranium-235 and -238(^{235}U, ^{238}U) and thorium-232 (^{232}Th). The decay of these isotopes proceeds sequentially through intermediate products having much shorter half-lives to the stable isotopes of lead, ^{207}Pb, ^{206}Pb, and ^{208}Pb, respectively. There are 14 other isotopes scattered through the periodic table that have half-lives spanning times from 10^{15} to 10^{9} years. All can be used for dating geologic materials. There are other isotopes, termed cosmogenic isotopes, that are produced by cosmic rays and that have half-lives that are useful for measuring periods of historical interest. The best known of these is carbon-14 (^{14}C), which has important uses in archaeology because of its half-life of 5760 years.

The age of the earth has been the subject of study since the beginning of human thought, and until the eighteenth century age of Enlightenment, theological and philosophical values were generally accepted, which for Western civilization gave an

age of about 10,000 years. Geologists reasoned from observations of sedimentary rocks and from the time required for the oceans to reach their present salinity. Based on the rate at which salt is delivered to the oceans by rivers, an age of many millions of years was determined. Toward the end of the nineteenth century, ages based on rather inappropriate models of the cooling of a hot primordial earth by Lord Kelvin disputed the ages preferred by geologists, but the discovery of radioactivity at the end of that century quickly altered the situation. Ernest Rutherford was the first to suggest the principle of using uranium to helium (U/He) ratios to compute the age of rocks. In 1906 he estimated from the amount of helium found in uranium-bearing rocks that they must be at least 500 million years old. Bertram Boltwood ascertained in 1907 that lead was the other final product of uranium decay and made an age estimate based on U/Pb ratios, but neither the accumulation of helium nor of lead yielded satisfactory ages for rocks when the method of determination was chemical (rather than isotopic) analysis. The existence of isotopes was not confirmed until 1913.

It was the development of mass spectrometry immediately after World War II that opened the way to accurate age determinations, as it then became possible to measure the concentration of a decay product relative to its parent isotope. Crucial to the technique is the separation of rubidium (^{87}Rb) and strontium (^{87}Sr) from one another with ion-exchange resin, as the mass spectrometer cannot distinguish the two mass-87 isotopes from one another. Consider the example of rubidium 87 decaying to strontium 87, perhaps the simplest in concept and one of the most important in application. Otto Hahn and colleagues proposed using the decay of rubidium for age determinations, but they were unable to do so because the mass spectrometer was not a tool in their experimental kit.

Age measurements followed a 1956 paper by L. Thomas Aldrich *et al.* in which he and his colleagues determined the laboratory techniques and decay constant. Although the rubidium is generally present in amounts too small for accurate chemical analysis, it can be determined by adding a known quantity of isotopically enriched rubidium, called the "spike," to the dissolved sample. The mass spectrometer yields rubidium isotope ratios from the mixture that allow one to determine simply and accurately how much natural rubidium is present. The measurements of strontium-87 are made relative to one of the other strontium isotopes, usually strontium-86, that are not of radiogenic origin. Comparison of the strontium-87 with the rubidium-87, allows the calculation of the age from the known rubidium-87 decay rate. An igneous rock to be dated is generally broken mechanically into a number of its mineral phases, each having a different concentration of the parent rubidium-87 and for which there is a measurement of the excess strontium-87. Presenting the data for each mineral graphically gives a plot, called an isochron, that must be a straight line if the data are consistent. Deviations from the line indicate that the rock has been subjected to heat, pressure, or liquid flow that has allowed the two chemically different atoms to migrate, invalidating the age determination.

Other isotope pairs with different geochemistry have application to dating of particular rocks. Using the decay of potassium (^{40}K) into argon (^{40}Ar) came at about the same time as Rb–Sr. Somewhat later, the decay of samarium (^{147}Sm) into neodymium (^{143}Nd) became as important as that of Rb–Sr. Because their refractory nature provided insight into certain geochemical systems, measurement of the decay of rhenium (^{187}Re) into osmium (^{187}Os) was long sought, but being refractory they could not be run on thermal ion source machines as atomic ions. This limitation came to a welcome end in 1989 when it was learned that these two elements form negative molecular oxygen ions that run very stably and accurately by the thermal method.

The thermal method utilizes the low ionization potentials of periodic table Groups 2a, 3b, and the lanthanides to effect ionization by temperature alone. The chemical element to be examined is extracted from the bulk sample and deposited on a ribbon-shaped filament, usually of tantalum, rhenium, or tungsten, which is placed in the mass spectrometer ion source. On heating to the temperatures attainable with such filaments, the deposited atoms are evaporated with many having been ionized. This technique, which has proved to be of great value in radioactive dating, has two advantages: (1) the ionization is selective, so that impurities that may have been deposited with the sample will not generally be ionized, and (2) the emitted ions have very little kinetic energy, specifically only that resulting from the temperature of the filament, which is of the order of 1 electron volt. Other methods of ionization yield distributions of energies of tens and hundreds of electron volts, requiring energy filters preceding the magnet.

Geologic ages rest on knowledge of the decaying parent isotopes, and the important isotopes of uranium, thorium, rubidium, and potassium have

half-lives in the range of 10^{10} and 10^{11} years. Except for uranium and thorium, which proceed to lead through the emission of a number of energetic and easily identifiable alpha particles, such long half-lives are difficult to measure in the laboratory. For this reason, very accurate measurements of uranium have been accepted as the standard. The other decay constants are determined by comparing age determinations of the same rock, meteorites being the best. Radiometric ages are generally believed accurate to better than 0.1 percent.

The use of cosmogenic isotopes differs from this simple procedure because there are no parent isotopes with which to compare them. The production of carbon-14 takes place in the upper atmosphere where it quickly bonds with oxygen to become carbon dioxide (CO_2) and is incorporated into photosynthesizing systems. After death there is no further carbon-14 replenishment, and it decays. To a first approximation the production rate is constant, so the ratio of carbon-14 to the stable carbon-12 can be used directly for the calculation of the age of the material containing the radioisotope. The rate of carbon-14 generation has varied over the past, so a correlation has to be made by studying a wide variety of archaeological materials of known age, especially tree rings of overlapping time spans. Radioactive counting of beta decays was initially the method for determining the amount of carbon-14 in a sample, but a special kind of mass spectrometry using nuclear particle accelerators has improved accuracy and extended the limits of maximum age determination.

The method is a variation of the magnetic sector type that uses a tandem Van de Graaff accelerator to measure isotope ratios of the order of 1×10^{-12}, ratios that are common in measuring cosmogenic isotopes. Such ratios are too small to be measured with conventional machines because the signal beam is "swamped" by ions scattered from the electrodes, chamber walls, and residual chamber gas. The accelerator furnishes ion energies a thousand times greater than in conventional spectrometers. Such high energies allow nuclear particle detector methods to identify individual ions of the cosmogenic atoms in the presence of huge background beams. The technique requires negative ions for the tandem Van de Graaff and by chance, five elements that have cosmogenic isotopes form such ions without interfering masses, although each requires a special experimental arrangement.

For uranium and thorium to be useful for dating, minerals must retain not only these elements but also their intermediate decay products and their lead. A mineral that best satisfies these requirements is zircon ($ZrSiO_4$), which has a strong affinity for uranium but excludes lead, and these characteristics have made it a workhorse for dating. In minerals in which the retention of uranium and lead differs sufficiently so that parent–daughter relationships cannot be used, researchers make use of uranium's unique property of having two isotopes with significantly different half-lives that give rise to two different isotopes of lead: uranium-235 to lead-207, and uranium-238 to lead-206. In the lead–lead method, the mass spectrometrist measures the lead isotopes relative to one another. Rocks of the same age that have lead-rich minerals that exclude uranium, such as galena (PbS), draw on sources with different uranium–lead compositions at the time of their hardening. For such a rock the lead isotopes taken from these minerals form an isochron that marks the time of the mineral's deposit. Claire Patterson measured the age of the earth in 1956 to be 4.55×10^9 years using the lead–lead method; this required determining the primordial isotopic composition of lead, the composition at the time of the earth's formation.

Procedures for measuring isotope composition have the disadvantage of destroying the extracted portion of the sample used. The phenomena that allow nondestructive analysis rely on exciting optical radiation or x-radiation from the atoms of the sample, most commonly induced by electron bombardment. In these methods, isotopic effects are too small to be accurately observed.

See also **Isotopic Analysis; Mass Spectrometry**

LOUIS BROWN

Further Reading

Aldrich, L. T., Wetherill, G.W., Tilton, G.R. and Davis, G.L. The half-life of ^{87}Rb. *Phys. Rev.*, 103, 1045–1047, 1956. Established the techniques of Rb/Sr dating.

Boltwood, B.B. On the ultimate disintegration products of the radioactive elements. part II: the disintegration products of uranium. *Am. J. Sci.*, 23, 77–88, 1907. Of historical importance.

Dalrymple, G.B. *The Age of the Earth*. Stanford University Press, Stanford, CA, 1991. A thorough history of the subject.

Harper, C.T., Ed. *Geochronology: Radiometric Dating of Rocks and Minerals*. Dowden, Hutchinson & Ross, Stroudsburg, PA, 1973. Reprints with comments on the basic papers.

Nier, A.O., Thompson, R.W. and Murphey, B.F. The isotopic constitution of lead and the measurement of geological time: III. *Phys. Rev.*, 60, 112–116, 1941.

Noller, J.S., Sowers, J.M. and Lettis, W.R., Eds. *Quaternary Geochronology: Methods and Applications*.

American Geophysical Union, Washington D.C., 2000. Among other things, a survey of radiocarbon dating but also many nonradioactive methods of dating.

Patterson, C. Age of meteorites and the earth. *Geochim. Cosmochim. Acta*, 10, 230–237, 1956. The age of the earth finally determined.

Wetherill, G.W. Discordant uranium–lead ages. *Trans. Am. Geophys. Union*, 37, 320–326, 1956. Mastering uranium–lead dating.

Radio-Frequency Electronics

Radio was originally conceived as a means for interpersonal communications, either person-to-person, or person-to-people, using analog waveforms containing either Morse code or actual sound. The use of radio frequencies (RF) designed to carry digital data in the form of binary code rather than voice and to replace physical wired connections between devices began in the 1970s, but the technology was not commercialized until the 1990s through digital cellular phone networks known as personal communications services (PCS) and an emerging group of wireless data network technologies just reaching commercial viability. The first of these is a so-called wireless personal area network (WPAN) technology known as Bluetooth. There are also two wireless local area networks (WLANs), generally grouped under the name Wi-Fi (wireless fidelity): (1) Wi-Fi, also known by its Institute of Electrical and Electronic Engineers (IEEE) designation 802.11b, and (2) Wi-Fi5 (802.11a).

Bluetooth replaces the physical connection between devices, such as between a personal desk assistant (PDA), digital camera, or printer to a personal computer (PC), or a cell phone headset with a cell phone. Bluetooth allows these devices to automatically interact; a PDA would automatically synchronize its database with its host PC whenever the devices come within range of each other. A Bluetooth WPAN operates on the 2.4-gigahertz-frequency band and allows devices to interact within a 10-meter area at 1 megabit per second (Mbps).

Wi-Fi (802.11b) is an Ethernet WLAN for the interconnection of PCs and mobile devices equipped with Wi-Fi transceivers, such as cable modems or other PCs, at 11 Mbps within 100 meters. The newer Wi-Fi5 (802.11a) operates at 54 Mbps within 100 meters. The propagation of individual Wi-Fi networks has resulted in larger wireless "hot spots" or wireless Internet service providers (ISPs). These anarchic hot spots theoretically allow an urban user to move from Wi-Fi node to Wi-Fi node for uninterrupted Internet access, in much the same way that a cell phone call is handed off from cell to cell.

Although the concept of wireless data networks is fairly new, the basic technology was created during World War II by two unlikely "scientists": Austrian-born actress Hedy Lamarr and orchestra leader George Antheil. Their U.S. patent described frequency-hopping, radio-controlled torpedoes that could not be jammed by the Nazis.

In the late 1950s, Sylvania's Electronic Systems Division devised an electronic version of frequency hopping and applied it to securing the military's satellite communications used during the 1962 Cuban Missile Crisis. Frequency hopping, now called spread spectrum, remained highly classified until the mid-1980s, by which time it had been developed for digital signal transmission, particularly for the code division multiple access (CDMA) protocol used by cell phones in North America and Asia.

In the late 1980s, the U.S. Federal Communications Commission (FCC) adopted rules for the commercial use of spread spectrum for low-power, unlicensed usage in the 902 to 928 megahertz, 2400 to 2483.5 megahertz, and 5752.5 to 5850 megahertz bands of the FM spectrum at a maximum of 1 watt peak output—the ISM (industrial, scientific and manufacturing) bands. Several European and Asian companies successfully lobbied their governments to adopt similar rules. Consumer telephone makers used these rules for longer-range cordless phones.

Equatorial Communications first commercialized spread spectrum data communications using a variation called direct sequence, which enabled access to multiple signals from synchronous satellites. Direct-sequence spread spectrum is used mostly in broadband applications such as Wi-Fi, while frequency hopping is used primarily for narrowband applications such as Bluetooth.

In 1997, the IEEE approved the first Wi-Fi specification, 802.11, which offered transmission speeds of 2 Mbps within 30 meters. Further improvements led in September 1999 to the adoption of Wi-Fi (802.11b) and Wi-Fi5 (802.11a). Using the wider 5-gigahertz bands instead of the 2.4-gigahertz bands, 802.11a accommodates more users, and its higher 54-mpbs speed enables multiple channels of video streaming. The 802.11g, expected to be adopted in early 2003, also offers 54 Mbps but uses the 2.4-gigahertz spectrum, making it compatible with 802.11b and enabling mixed Wi-Fi networks.

In 1994, Ericsson began research on Bluetooth, a narrowband WPAN frequency-hopping spread

spectrum system named after a tenth century Danish king, Harald Blaatand (Bluetooth). In May 1998, Ericsson, IBM, Motorola, Intel, and Toshiba formed the Bluetooth Special Interest Group (SIG). In 1999, 3Com, Lucent, Microsoft, and Motorola joined the group, and the first Bluetooth specifications were published. These included specifications for a broad array of applications, or "profiles," which include wireless interconnection between stereo audio components and speakers.

In the summer of 1999, the IEEE began efforts to standardize Bluetooth specifications. On 21 March 2002, the IEEE 802.15 Working Group for Wireless Personal Area Networks approved IEEE 802.15.1, which is based on Bluetooth SIG specification version 1.1.

While Bluetooth and Wi-Fi are designed to provide clean and secure connections within the 2.4-gigahertzband, the two technologies can interfere with one another. The IEEE 802.15.2 specification, which was scheduled for adoption sometime in 2003, is designed to facilitate coexistence between the two. The Bluetooth SIG also is considering Wi-Fi coexistence.

In February 2002, the FCC approved limited deployment of a faster and broader wireless pipeline called ultrawideband (UWB), which could operate anywhere between 3.1 to 6 gigahertz at speeds of 100 to 500 Mbps within 10 meters. UWB was initially approved for applications such as collision avoidance, ground-penetrating radars, and wall-penetrating imaging systems, but questions persist about potential interference with other wireless technologies, such as global positioning satellite (GPS) systems. The IEEE is exploring the integration of UWB into 802.15.3, also called WiMedia, offering 110 Mbps in the 2.4-gigahertz band for use in imaging and multimedia applications.

There are a number of other wireless data transmission standards on the drawing boards, including IEEE 802.16, a metropolitan area network (MAN) that features 10 gigabits per second (Gbps) transmission speeds. Several worldwide spectrum regulatory bodies also have set aside spectrum in the 5-gigahertz band designed for high-speed Internet access in industrial applications. At the time of writing, only a handful of devices designed to take advantage of these bands has been announced.

See also **Electronic Communications; Mobile (Cell) Telephones; Satellites, Communications**

STEWART WOLPIN

Further Reading

Garretson, C. FCC leaning toward deregulated broadband. *InfoWorld*, 14 February, 2002.
Harmon, A. Good (or unwitting) neighbors make for good Internet access. *The New York Times*, 4 March, 2002.
Lansford, J. Can Bluetooth, Wi-Fi coexist in the future? *EE Times*, 1 February, 2002.
Leopold, G. FCC gives cautious nod to ultra-wideband. *EE Times*, 14 February, 2002.
Mannion, P. IEEE approved Bluetooth-based wireless PAN standard. *EE Times*, 22 March, 2002.
Markoff, J. The corner Internet network vs. the cellular giants. *The New York Times*, 3 March, 2002.
Reuters. FCC okays limited use of ultrawideband technology. *InfoWorld*, 14 February, 2002.
Vaughan-Nichols, S. Too many wireless LAN standards. *Enterprise*, 11 December, 2001.
Yoshida, J. UWB MACs: build or borrow? *EE Times*, 1 February, 2002.
Yoshida, J. and Leopold, G. Ultrawideband companies gear up for FCC ruling. *EE Times*, 1 February, 2002.
Zyren, J. 802.11g spec: covering the basics. *EE Times*, 1 February, 2002.

Useful Websites

Bluetooth, The Official Bluetooth Web site: http://www.bluetooth.com/
Hytha, D. Bourk, T., Linsky, J. Bluetooth adaptive hopping: tech challenges and regulatory headaches. *Comms. Design*, 13 February, 2002: http://www.commsdesign.com/story/OEG20020213S0015
IEEE 802.11 Wireless Local Area Networks, Working Group: http://ieee802.org/11/
IEEE 802.15 WPAN Task Group 1 (TG1): http://www.ieee802.org/15/pub/TG1.html
Siep, T.M., Gifford, I.C., Braley, R.C., and Heile, R.F. Paving the way for personal area network standards: an overview of the IEEE P802.15 Working Group for Personal Area Networks. *IEEE Personal Communications*, February 2000: http://grouper.ieee.org/groups/802/15/pub/2000/Mar00/99012r6P802-15_WG -PCM-Article.pdf
TI brings Wi-Fi (802.11b) to the masses. Texas Instruments press backgrounder: http://www.ti.com/corp/docs/press/backgrounder/80211.shtml
Wireless Ethernet Compatibility Alliance (WECA) website: www.wi-fi.org

Radionavigation

Astronomical and dead-reckoning techniques furnished the methods of navigating ships until the twentieth century, when exploitation of radio waves, coupled with electronics, met the needs of aircraft with their fast speeds, but also transformed all navigational techniques. The application of radio to dead reckoning has allowed vessels to determine their positions in all weather by direction finding (known as radio direction finding, or RDF) or by hyperbolic systems, discussed below.

Another use of radio, radar (radio direction and rangefinding), enables vessels to determine their distance to, or their bearing from, objects of known position. Radionavigation complements traditional navigational methods by employing three frames of reference. First, radio enables a vessel to navigate by lines of bearing to shore transmitters (the most common use of radio). This is directly analogous to the use of lighthouses for bearings. Second, shore stations may take radio bearings of craft and relay to them computed positions. Third, radio beacons provide aircraft or ships with signals that function as true compasses.

Radio developments accelerated before and during World War I. The *Mauretania* was the first ship to carry a radio receiver for taking bearings (1911). At the Battle of Jutland, Royal Navy ships tracked German fleet movements by radio. However, it was not until after 1921, when the first land-based transmitters were installed for continuous navigational use, that RDF enabled aircraft and ships to take bearings in any weather without dependence on actual sightings. In fact, bearings can be taken from virtually any detectable transmission source. The earliest receiver found on aircraft was the loop antenna, developed by the Marconi Company, also called a "radio compass." The loop rotated electronically to detect a transmission and enable a bearing to be taken (see Figure 23). Figure 24 represents a common pre-World War II short-wave RDF receiver employing antennas at right angles. Similar systems are still used, particularly common in automatic direction-finding equipment. Radionavigation, despite its advantages, is affected by atmospheric refraction, attenuation, polarization, quadrantal error (reflection from the craft itself), and differences between a visual line of bearing and one obtained by radio (see hyperbolic systems below). Furthermore, frequencies also vary in range, ionospheric distortion, and accuracy. By the 1930s, mid- to high-frequency ranges were employed for most radio applications.

World War II created an industry for improved methods of radionavigation for night bombing. The Royal Air Force deployed very high frequency (VHF) RDF stations for this purpose, while the Germans installed radiobeacons that broadcasted distinctive pulses along compass points to direct bombers to British targets.

Hyperbolic systems developed to meet tactical needs. In a hyperbolic system, a master transmitter broadcasts a distinctive series of pulses that is repeated at intervals by two slave stations. A ship, receiving signals from all transmitters, measures the time differences in their arrival at the receiver

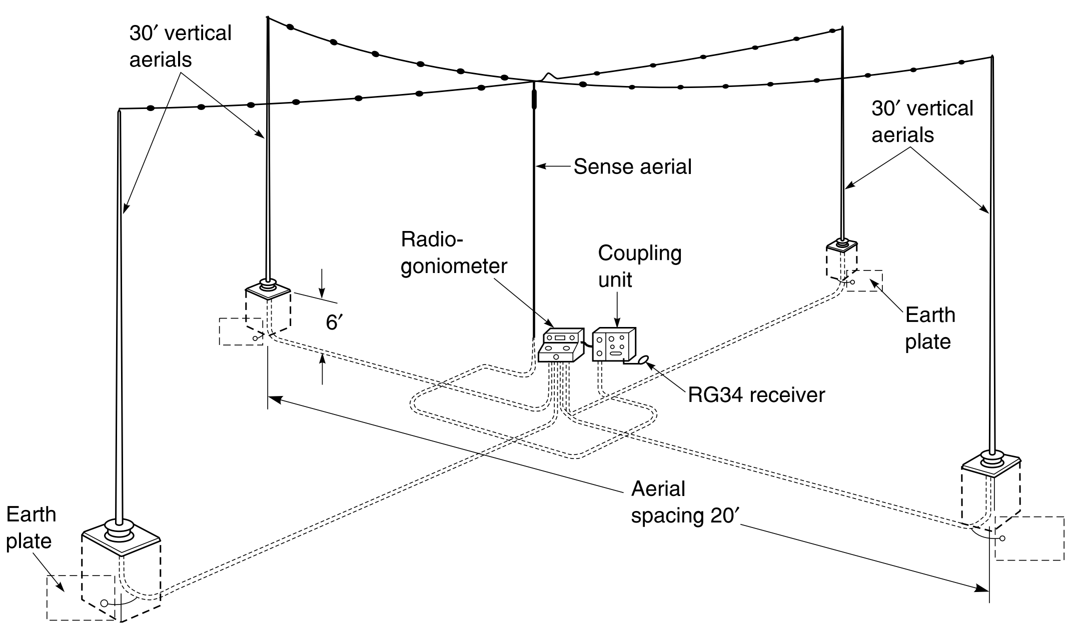


Figure 23. A loop antenna commonly found on aircraft before and during World War II. The loop couples to an autosyn that gives a relative bearing to the transmitting source. The loop aligns to a signal either manually or electrically.
[*Source: U.S. Navy. Air Navigation Part Four: Navigation Instruments, Flight Preparation Training Series. McGraw-Hill, New York, 1944, p. 103.*]

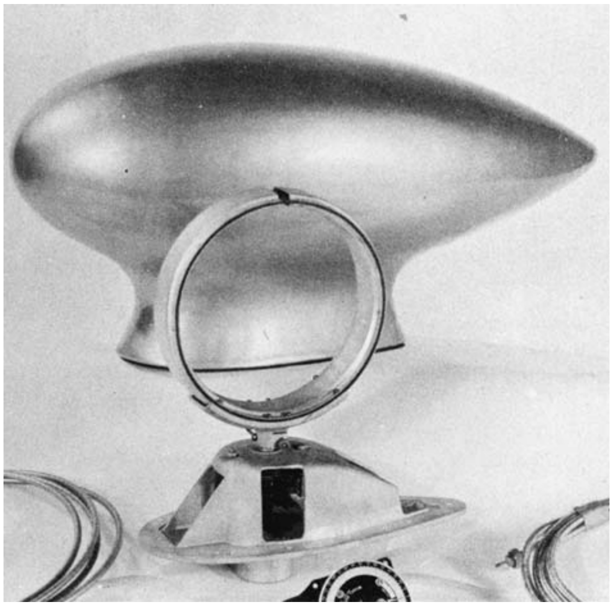


Figure 24. Marconi–Adcock RDF equipment from the 1930s. The goniometer determines the direction of signals received.
[*Source: Weems, P.V.H. Air Navigation, 2nd edn. McGraw-Hill, New York, 1938, p. 225.*]

to determine distance from each station. On a chart or plot, points at which signals show the same time difference define a hyperbola, a line of position. The points of intersection of multiple hyperbolas define a fix.

By 1942, the U.K. had established the first hyperbolic system, Gee, operating between 30 and 80 megahertz. Before the war, the Germans developed their hyperbolic system, Sonne, later adapted as Consol by the Royal Air Force. The U.S. followed closely with Loran (long-range navigation) in 1943. This method, later called Loran-A, developed after the war into Loran-C, which covers 1300 to 2300 kilometers through broadcast at 90 to 110 kilohertz, and although considered obsolete by world navies by the end of the century, remains the most successful and widely used marine radionavigation system. A British hyperbolic system, Decca, operating between 70 and 130 kilohertz (with a range under 500 kilometers), relies not on the timing of pulsed signals, but instead compares phase shifts between signals from multiple transmitters against a reference signal to determine location. First used under the name QM to support the D-Day invasion in 1944, Decca remains in use for coastal navigation, especially commercial fishing. The last hyperbolic application of the century was Omega, (still in global use), which operates between 10 and 14 kilohertz (with a range of up to a few thousand kilometers) and, like Decca, measures phase shifts.

Various researchers and inventors theorized that the Doppler shift of radio waves along the spectrum owing to motion toward or away from the measuring device had a navigational application. Following the launch of Sputnik-1, researchers at the Applied Physics Laboratory, Johns Hopkins University, developed Doppler techniques (already in aircraft navigational use) to track satellite orbits and determined that if a satellite's orbit was known, then the measurement of the satellite signal's Doppler shift with respect to an earthbound point could determine location. The U.S. Navy immediately applied the principle to the development of a satellite-based navigational system known as NAVSAT or TRANSIT, which became operational in 1964. Operating at 400 megahertz, TRANSIT satellites occupy polar geosynchronous orbits 1000 kilometers from the earth and deliver an accuracy of 0.1 nautical mile (around 185 meters). TRANSIT, however, applies mainly to slow-moving ships, and began to be phased out by the late 1990s, when it was superseded by the global positioning system, or GPS. GPS became operational in 1995 with 24 satellites in six orbital planes. In both military and commercial applications, GPS can provide time, location, and speed data to any user. A user's receiver selects signals from four satellites and synchronizes them to obtain locational data based on a comparison of time signals emanating from clocks internal to the satellites and receiver. With the launch of additional satellites, GPS promises reliability and accuracy, and will doubtless supersede other navigational methods.

Radar followed a concurrent development with RDF and hyperbolic navigation. During the 1920s, the detection of ships by reflected radio waves had been demonstrated, and by the outset of World War II, radar detection was established in the U.K. Radar equipment measured a time interval between the transmission of a signal and its reception as an echo as reflected from an object. Radar had been installed on naval ships during the 1930s, but its accuracy was poor. British scientists at Birmingham University pioneered radar's wartime breakthrough with the invention of the multicavity magnetron, which permitted the Royal Navy to deploy imaging at centimeter wavelengths, a huge improvement in accuracy. British authorities shared the invention with the U.S., where scientists developed microwave radar during the war. Following the war, radar saw wide navigational applications, and by the end of the

century, most vessels and aircraft were required to carry radar.

Aircraft and ships today receive continuous positional data through radio fixing. Ships rely on radionavigation in conjunction with other methods, while aircraft have been more radio-dependent. Submarines and missiles rely on inertial methods, although the advent of GPS may eventually replace all other methods of radionavigation.

See also **Global Positioning System (GPS); Gyrocompass and Inertial Navigation.**

ROBERT D. HICKS

Further Reading

Defense Mapping Agency Hydrographic/Topographic Center. *The American Practical Navigator*, Pub. No. 9, Bethesda, MD, 1995.
May, W.E., *A History of Marine Navigation*. W.W. Norton, New York, 1973.
Page, R.M. *The Origin of Radar*. Anchor, New York, 1962.
Sonnenberg, G.J. *Radar and Electronic Navigation*, 6th edn. Butterworth-Heinemann, London, 1988.
Tetley, L. and Calcutt,D. *Electronic Navigation Systems*, 3rd edn. Butterworth-Heinemann, London, 2001.
Williams, J.E.D. *From Sails to Satellites: The Origin and Development of Navigational Science*. Oxford University Press, Oxford, 1994.
Wylie, F.J. *The Use of Radar at Sea*. Hollis & Carter, London, 1952.

Rail, Diesel, and Diesel-Electric Locomotives

Gasoline and oil-engine trolleycars appeared in the late nineteenth century. Internal combustion locomotives were first tried in dockyards, on military and contractor's railways, and where steam traction was undesirable, but these were of low power and limited use. Diesel traction owed much to the thermal electric principle demonstrated by J.J. Heilmann in the 1890s using locomotives that carried their own electricity-generating steam plant. Between 1900 and 1920 the oil engine was put on rail: passenger vehicles and baggage cars were equipped with gasoline or oil motors for branch line work; road vehicles were fitted with rail wheels; and a few new intercity trains were constructed. Oil-engine locomotives were, however, relatively rare.

Pioneering work in Scandinavia around 1912, Germany in 1914, and the U.S. about 1917 was crucial. Direct mechanical drive to the wheels via shafts and gear boxes, sometimes with hydraulic drive, was suitable for low or moderate power, and electric transmission was favored for greater power, especially in the U.S. General use of the diesel engine with electric transmission, was retarded by the lack of a system to control engine, generator, and traction motors working together. This was eventually provided by Lemp in the U.S.

Electric and diesel-electric locomotives used common components, and multimode units were constructed in the 1920s with a diesel engine, batteries, current collection shoes, and pantograph. However, a considerable increase in power-to-weight ratio was needed for the diesel locomotive to become a general traction unit. The Winton Engine Company, engine suppliers to the General Motors-owned Electro-Motive Company, produced the motor needed. Harold Hamilton and Richard Dilworth founded the Electro-Motive Company and built over 500 rail cars between 1924 and 1933. Electro-Motive was taken over by General Motors (GM) in 1930 but remained an autonomous concern. During the interwar period, General Electric (GE) led diesel locomotive work and produced several noteworthy demonstration units but lost the lead to General Motors in the 1930s.

The U.S. led the dieselization of railways after 1930. Working relationships were established between steam locomotive builders (Alco, Baldwin, and Lima), the big electrical combines such as Westinghouse and GE, and the makers of internal combustion prime movers such as Ingersoll and GM. Development projects were controlled by large companies that had the range of engineering skills and funds. In the U.S., electric transmission was standard from the beginning, because low-voltage direct current (DC) systems dominated railway traction and street trolleys. The DC motors were ideal for railway service, and were therefore used for the diesel locomotive in the U.S.

In Europe the suitability of electric transmission was questioned. Many forms of transmission were tried, including steam, air, exhaust gas, and mechanical systems. These "compound diesels" demanded radical departures from the orthodox form of oil engine. Hybrid engines, like the Still, combined the diesel cycle with steam using an exhaust-heated boiler, but the system failed to become established. In Russia, Lenin funded investigations by G.W. Lomonossoff into the main types of transmission, with considerable German help. He found that electric transmission was the best method of transmitting work from the engine to the wheels over a wide speed range. In Russia, there was considerable theoretical analysis of locomotive thermodynamics and design, and one of the first university chairs devoted to the diesel locomotive was set up in Moscow. In

Germany there was close cooperation between the railway company's research departments, technical university specialist schools, and the manufacturers' research teams. However, no European establishment matched the organized program mounted by GM to dieselize railways. Several lightweight, high-speed passenger trains powered by diesel engines were demonstrated in the 1930s, but these were not powerful enough to displace the best steam locomotives of the time.

GM used automobile industry lessons of manufacture, marketing, publicity, sales technique, and training and applied them in launching diesel traction. Production designs were standard, hire-purchase terms encouraged railways to accept diesels, demonstration units were offered on loan to work prestige train routes. In 1935, two 1800-horsepower Electro-Motive units coupled together worked the Sante Fe Super Chief straight through from Chicago to Los Angeles. Free training of engine crew was available, and freight diesels toured North America demonstrating their superiority to steam traction. Although, dependent at first on GE for manufacture of electrical parts, GM set up its own electrical factory to meet its locomotive needs. Independent after 1938, GM built up a lead that other manufacturers could not overtake. GE failed to displace the GM diesel with a steam-electric locomotive called the Steamotive, but using its considerable resources was able to capture a fair share of the market. Other builders in the U.S. tried to introduce diesels of their own, but went out of business in the 1950s.

Following World War II, postwar recovery led to dieselizing of nonelectrified railways in Europe. Experiments with gas-turbine-electric locomotives and advanced diesel-mechanical and hydraulic locomotives failed to displace the diesel-electric as the best complement to electric traction. In the U.S., diesel-electric traction was regarded as "cheap electrification" that avoided fixed works.

In Europe, the diesel engine became a power unit in all forms of train. The pioneer work of Franz Kruckenberg in Germany in the 1930s, led to fast diesel trains in the postwar Benelux countries, continuing development that had started with a train called the Flying Hamburger. By the 1960s, British high-speed trains showed that diesels could work long-distance express services faster than steam and could rival contemporary electric operations. Present day electric *Train á Grande Vitesse* (TGV) operations far exceed the speed of any diesel service, but the modern diesel train can regularly work at 250 kilometers per hour. In the 1930s diesel began to displace steam and electricity in freight hauling. In the U.S. and Russia, diesels run in multiple have moved test trains of over 100,000 tons in weight controlled from one cab.

Electronics has transformed the diesel locomotive. Component size has been reduced. Solid-state rectifiers enabled the DC generator to be replaced by the smaller alternating current (AC) alternator. The solid-state inverter and control apparatus enabled the asynchronous motor to replace the traditional DC traction motor. Globalization of industries in the 1980s led to designs for a universal diesel locomotive able to work anywhere in the world. This philosophy gave way to the modular policy of building diesel and electric units, from heavy goods engines to TGV, using common components. At the close of the twentieth century, European companies cooperated with North American combines to produce global designs, and GM diesel-electric locomotives made in North America were at work in Europe. The diesel locomotive continued to develop: driverless working was demonstrated; and experiments to operate diesel locomotives with hydrogen, coal-based fuels, methane and other gases, and coal-slurry have been carried out. Though it cannot match the very high performance of the electric locomotive, the diesel is ideal for all operations outside electrified zones.

See also **Rail, High Speed**

MICHAEL C. DUFFY

Further Reading

Brown, H.F. Economic results of diesel-electric motive power on the railways of the United States of America. *Proc. Inst. Mech. Eng.*, 175, 5, pp 257–317, 1961.

Cummins, L. *Diesel's Engine, Vol.1: From Conception to 1918*. Carnot Press, Oregon, 1993.

Duffy, M.C. The Still engine and railway traction," *Trans. Newcomen Soc.*, 59, pp 31–59, 1987–1988.

Dunlop, J. Internal combustion locomotives. *Trans. Inst. Eng. Ship. Scotland*, 68, 234–278, 1925.

Fell, L.F.R. The compression ignition engine and its applicability to British railways. *Proc. Inst. Mech. Eng.*, 124, 1–66, 1933.

Hinde, D.W. *Electric and Diesel Locomotives*. Macmillan, London, 1948.

Kirkland, J.F. *Dawn of the Diesel Age*. Interurban Press, Glendale, CA, 1983.

Morgan, D.P. Diesel traction in North America, in *Concise Encyclopedia of World Railway Locomotives*. Hutchinson, 1955, pp 107–142.

Tufnell, R.M. *The Diesel Impact on British Rail*. Mechanical Engineering Press, 1979.

Vickers, R.L. The beginnings of diesel-electric traction," *Trans. Newcomen Soc.*, 71, 1, 115–127, 1999–2000.

Rail, Electric Locomotives

Electric locomotion grew from nineteenth century electrification of quarries, mines, light railways, and cable car systems, and basic forms were established in the street trolleys of the 1880s in the U.S. Heavy-duty electric traction evolved in the 1890s on rapid transit railways in large cities like Chicago and New York. Low-voltage direct current (LVDC) technology, using rotary converters and conductor rails, was needed for terminal and tunnel lines, where locomotives were required. Locomotives equal in power to any steam locomotive were working in the period from 1900–1910. Overall energy efficiency was low if power came from coal stations and electric traction was employed to meet legal requirements or operational needs. Early locomotives took many forms: some used body-mounted motors driving through rods and shafts; others used geared axle-mounted nose-suspended motors. Other than heavy-duty rapid-transit railways, electric traction could not replace steam traction on general-purpose railways. The Pennsylvania Railroad and New York Central terminal lines and the Detroit and Baltimore tunnel lines were early locomotive worked railways.

From 1900 to 1920, other electric transport systems appeared, including cableways (Germany, Switzerland), monorails (Langen, Kearney), and funiculars (Italy). Electric rail "mules" were used to haul ships through the Panama Canal locks.

The rail-less trolleybus, locomotive, and train, which drew power from an overhead line and ran on road wheels, appeared before World War I (1914–1918). Road locomotives powered from batteries found use. A hybrid vehicle, with engine, generator, and accumulators, was invented, and found use throughout the twentieth century in passenger transport over routes too lightly used to justify electric tramways. Heavy dumper trucks that draw power from an overhead supply line continue to be of use in quarries. The use of electric rail-less road vehicles declined with the rise of motor transport. Many electric trolley systems were dismantled, but there has been a general revival of electric railways since 1970.

The early heavy-duty electric railways, with third-rail direct current (DC) distribution were not suitable for long-distance mainline work. The three-phase railway dated from the Lugarno tramway (1896), and was extensively used in Italy, on the Simplon Tunnel route (Switzerland), and through the Cascade Tunnel (U.S.). The need for two overhead wires and speed-control limitations prevented widespread use. By the early 1930s this system had been replaced by monophase systems in Switzerland and the U.S., although it survived in Italy until after World War II. The single-phase alternating current (AC) railway dates from 1902 with the introduction of a reliable, powerful single-phase commutator motor. The 1905 tests of 11-kilovolt single-phase AC locomotives on the Pennsylvania Railroad in 1905 established the system, and in 1907 it was used on the New Haven Railroad lines into New York. In 1914 it was installed on the Philadelphia Paoli Division of the Pennsylvania Railroad.

In Europe, the 1908 Loetschberg line in Switzerland used 15 kilovolts and 16.66 hertz, which became standard in Austria, Switzerland, Germany, Sweden, and Norway until replaced by the 25 kilovolt, 50 hertz European standard in the 1950s. Tests on the Seebach–Wettingen line between 1901 and 1905 with phase converter locomotives investigated various motor and control systems drawing power from a single contact wire. The first locomotive to use industrial frequency (50 hertz), at 15 kilovolts, was tested there.

By 1913, locomotives of 2500 horsepower were in use. AC motors of great power were large and often body-mounted, with rod and jackshaft drive and many engineers favored the DC motor. Converter locomotives used single-phase AC supply with DC motors by having an onboard rotary converter. These were also heavy, large machines. The first in the U.S. was constructed in 1925 by Henry Ford at the Rouge River Plant for the Ford-owned Detroit, Toledo and Ironton railroads. Its supply was 11 kilovolts and 25 hertz; its traction power was 3100 kilowatts; and its weight was 393 tons.

Converter locomotives were never numerous. The final American examples were supplied in 1948 to the Great Northern Railway (320 tons, 5000 horsepower) and to the Virginian Railway (454 tons, 7800 horsepower). It was rendered obsolete by the locomotive mercury-arc rectifier, and the solid-state rectifier.

The Hungarian Kando experimented between 1927 and 1941 to combine single-phase AC supply with DC motors, but size, weight and complexity prevented general use. By the mid-1930s rectifier locomotives using mercury vapor were under trial. Phase-splitting locomotives, which converted single-phase AC to three-phase supply to induction-type traction motors, were used on the Norfolk and Western Railway in 1915 and on the Virginian Railway. Its supply was single-phase 11 kilovolts and 25 hertz.

The high-voltage direct current (HVDC) railway generally used a supply voltage of 1500 or 3000

volts. This was proven in 1911 on the Butte, Anaconda, and Pacific Railway in Montana, and used on the Montana Division electrification scheme (700 kilometers) of the Chicago, Milwaukee, and St. Paul Railway, completed in 1915. This showed that a general-purpose steam railway could be better worked by electric traction. All sections used 3000 volts DC. These installations influenced strategy worldwide, and HVDC became a world standard.

The spread of railway electrification was encouraged by the construction of high-efficiency thermal power stations, the use of the mercury-arc rectifier as a static rectifier and phase converter, and the standardization of grid supply after 1930. In Europe, electrification was aided by national ownership of railways The wars, economic difficulties, and political crises of Europe and Asia between 1914 and 1950 delayed many electrification projects. In the U.S., the diesel-electric locomotive met the needs of general-purpose railways, and electric locomotive development ceased until the 1980s. By the late 1930s, electric locomotives were outperforming steam traction in every field. In the 1950s there were increasing demands for much higher speeds, well in excess of 160 kilometers per hour, and the superiority of electric traction became evident. The ability of electric locomotives to integrate closely with electric signaling and control technologies was advantageous.

After World War II, the 25-kilovolt 50-hertz single-phase system became the world standard. High-speed trials in France, under the 1500-volt DC system, set a world speed record of 331 kilometers per hour with locomotive BB9004 in March 1955. This record stood until 1981 when a SNCF *Train á Grande Vitesse* (TGV) train reached 371 kilometers per hour. The locomotive-mounted mercury-arc rectifier combined the single-phase AC supply with DC traction motors. Hibbert conducted pioneer experiments in 1913, but regular use did not begin until 1950. Locomotive mercury rectifiers proved unreliable and were replaced by solid-state devices from the mid-1960s. The success of solid-state rectifiers, inverters, and controllers enabled any type of supply system to be used with any kind of traction motor. In the 1970s Brown–Boveri developed variable speed control for modern three-phase traction motors. Solid-state devices provided multistage transformations of energy between supply and motors with minimal loss. They introduced reliable, compact technologies for suppressing interference between traction equipment and signaling and control networks. Onboard information processing systems shifted trackside signaling into the locomotive and made possible the automated, driverless mainline railway. Driverless railways were previously simple systems like the London Post Office Railway of the 1920s. Rapid transit lines in San Francisco, London, and other large cities were automated after the 1960s. Mine railways were worked without drivers in the U.S., and driverless goods trains tested on main lines in Germany. By the close of the twentieth century it was possible to monitor all locomotive operations from a control satellite in geostationary orbit. Automatic control of all operations is likely to be introduced in the early twenty-first century.

In the 1970s the "universal" electric locomotive was designed for most duties. The DB Class 120, introduced in 1979, is one example. The "modular" philosophy became dominant, and a range of locomotives, from shunting engines to very high-speed trains use components in common. Design has been "globalized" by the formation of worldwide combines. European-based industries such as Adtranz and Alstom are moving toward partnership with American companies like General Electric and General Motors. In Great Britain, privatization of the nationalized railways has slowed new electrification because private operators find diesel traction cheaper. In North America, the diesel remains the norm, despite occasional proposals to use dual-mode locomotives over partly electrified routes. On the high-speed railways found in most industrialized countries, both multiple-unit and locomotive-hauled trains generally run at speeds of 350 kilometers per hour, and speeds of 515 kilometers per hour have been attained. The limits of wheel-on-rail transport have not yet been reached, and conventional TGV trains have outperformed maglev (magnetic levitation), jet-trains, tracked hovercraft, and monorails. Apart from short run passenger conveyers, long-distance maglev routes are unlikely, despite the engineering success of the German and Japanese test installations. In terms of passenger-carrying capacity, loads moved, and speed, the electric railway worked with conventional locomotives and multiple units has much unrealized potential. The effects on railways of applying electronic control and guidance to road vehicles and modern road-trains (the "automated highways" concept) have yet to be determined.

See also **Rail, Diesel and Diesel Electric Locomotives; Rail, High Speed; Rail, Steam Locomotives; Railway Mechanics**

MICHAEL C. DUFFY

Further Reading

Agnew, W.A. *Electric Trains: Their Equipment and Operation*, 2 vols. Virtue, London, 1937.

Bailey, M.R. The tracked hovercraft. *Trans. Newcomen Soc.*, 65, 129–145, 1993–1994.

Bezilla, M. *Electric Traction on the Pennsylvania Railroad, 1895–1968*. Pennsylvania State University Press, 1980.

Binney, E.A. *Electric Traction Engineering*. Cleaver Hume, 1965.

Carter, F.W. *Railway Electric Traction*. Arnold, 1922.

Dover, A.T. *Electric Traction*. Pitman, London, 1954, 1929, 1918, and 1917. History of electric traction theory and practice traced through successive editions of standard text.

Duffy, M.C., *Electric Railways 1880–1990*. Institution of Electrical Engineers, London, 2003. History of progressive integration of traction, power supply, signaling, and control elements of electric railway system.

Haut, F.J.G. *The History of the Electric Locomotive*. George Allen & Unwin, 1969.

Hinde, D.W. *Electric and Diesel Locomotives*. Macmillan, 1948.

Johnson, J. and Long, R.A. *British Railways Engineering 1948–1980*. Mechanical Engineering Publications, 1981. Contains review of British electrification policy and electric traction equipment.

Kaller, R. and Allenbach, J.-M. *Traction Electrique*, 2 vols. Presses Polytechniques et Universitaires Romandes, Lausanne, 1995. Review of current railway and tramway practice.

Machefert-Tassin, Y.M., Nouvion, F and Woimant, J. *Histoire de la Traction Electrique, Tome 1: Des Origines a 1940*. La Vie du Rail, Paris, 1980; *Tome 2: De 1940 a Nos Jours*, La Vie du Rail, Paris, 1986. The most thorough and authoritative history of electric traction yet available.

Manson, A.J. *Railroad Electrification and the Electric Locomotive*. Simmons Boardman, New York, 1925.

Middleton, W.D. *The Interurban Era*. Kalmbach, 1961.

Middleton, W.D. *The Time of the Trolley*. Kalmbach, 1967.

Middleton, W.D. When the Steam Railroads Electrified. Kalmbach, 1974.

Raven, V. Railway electrification. *Trans. North-Eastern Coast. Eng. Shipbuild.*, 38, 173–240, 1921–1922.

Raven, V. Electric locomotives. *Proc. Inst. Mech. Eng.* Paris Meeting issue, 1922, pp. 735–781, 1057–1082.

Semmens, P and Machefert-Tassin, Y. *Channel Tunnel Trains: Channel Tunnel Rolling Stock and the Eurotunnel System*. Eurotunnel, 1994. Review of equipment and operations of modern electric railway.

Rail, High Speed

High speed railway operations in the twentieth century resulted from deliberate attempts to work a system at speeds much higher than the maximum speeds attained on conventional railways, with competition for the shortest time on routes between large cities; for example, passenger train rivalry on routes between New York and Chicago in the 1930s. High-speed railways required construction of new lines; segregation of operations; new forms of rolling stock, and unorthodox methods of propulsion relying on techniques developed outside the railway industry. However, the great success of the segregated high-speed railway is a post-1960s phenomenon.

In the 1890s the "crack" express trains might have reached 145 kilometers per hour (km/h) for short periods, but sustained working at 95 km/h (or 60 miles per hour) was considered adequate. A first-class railway could be proud of its "mile-a-minute" express timetables, start-to-stop. It was a wiser objective to increase the number of express trains averaging 95 km/h, and briefly touching a maximum of 145 km/h, than to seek ways of running at speeds far in excess of 150 km/h. In the years from 1900 to 1914, a few very fast trains were run by companies in Europe and the U.S. using conventional rolling stock, which may have exceeded 160 km/h during unscheduled attempts at a record. These reports were not authenticated and remain doubtful. Scheduled running in excess of 160 km/h tested contemporary technology and made little economic sense.

In the closing years of the nineteenth century, some experiments were made to see how fast trains could run. There were various schemes for constructing completely new railways, powered by electricity, and intended to run streamlined expresses at speeds well in excess of 200 km/h. Many of these schemes were fraudulent with a main aim of robbing unwise investors of their money, and few made engineering sense. One early project that displayed engineering skill if not economic acumen was due to David G. Weems, an American dentist who built a narrow-gauge electric railway in 1889 to attract funds for a full-sized system. A model car of some 2 meters in length reached a speed of 192 km/h over a 2-mile (3.2-kilometer) circuit. It was driven by electricity picked up from overhead wires with return through the rails.

In some quarters the steam-electric locomotives of J.J. Heilmann, built in the 1890s, were seen as potential motive power units for high-speed railways, despite their intended role in the electrification of ordinary railways. Radically different systems intended to replace steam railways were proposed by competent engineers driven by a utopian vision of rapid communication fostering universal brotherhood. Some concepts failed technically; others were of limited use. None of the proposed alternatives was particularly fast.

Serious, scientific engineering research into very high-speed railways began with investigations in Germany by Siemens & Halske and by AEG

between 1899 and 1903. These trials demonstrated the potential of electric traction in sustained running at speeds that no steam locomotive could match. The trials had state backing and took place on a military railway between Marienfeld and Zossen, near Berlin. Three vehicles were tried: one locomotive not intended for high-speed working, and two motorized carriages which were to set long-standing records. The vehicles were fitted with three-phase motors, with a 10 kilovolt supply picked up by a triple collector from three overhead contact wires. Speed was controlled by varying the speed of the steam engine driving the alternator according to signals telegraphed from the cars to the power house, which varied frequency: either 25 or 50 hertz. Speeds of the order of 210 km/h were reached by both cars. An attempt to match this performance in 1904 by a specially constructed 4-4-4 steam locomotive failed, though the design, by Kuhn, was to influence later German attempts to work steam railways at very high speeds. There was no compelling demand between 1900 and 1950 for very fast express passenger trains able to run above the maximum set by existing engineering limits and the need to work many kinds of trains, at very different speeds, over a common system. The years of war, economic difficulty, political uncertainty, postwar reconstruction of basic industries, and reorganization of many national railways, were not favorable to high-speed railways. The emphasis was on improving, or making do, with what existed. Some extremely valuable research into fast running and vehicle design was undertaken by Franz Kruckenberg in Germany in the 1930s. In Europe and the U.S. improved regular services between major cities were provided by short, lightweight trains powered by internal-combustion engines. The Zephyrs in the U.S., and the Flying Hamburger and similar trains in Germany and the Netherlands were examples. For high-speed operating of heavy trains over several hundred kilometers, only steam traction was available in the 1930s. The Chicago North Western's Hiawatha trains, the British LMS Coronation Scot, and LNER Coronation and Silver Jubilee demonstrated steam traction at a peak that would be difficult to surpass. These services severely taxed existing signaling and made the routing of slower trains difficult. Record-breaking runs with steam locomotives, like the LNER Mallard, which reached 203 km/h, questioned the wisdom of trying to work trains by steam at sustained speeds of over 160 km/h.

During World War II, a detailed study of high-speed railways was undertaken by the German state in the years when it was confident of victory. Plans were made for a network of 250 km/h services throughout the Greater German Reich and its conquered territories, some using improved standard-gauge routes and others using a proposed broad-gauge system promoted by Hitler. Very thorough studies were completed, and the broad-gauge project continued until halted by the defeat of Germany in 1945. Many modes of traction were considered, but electric traction was not favored. The standard-gauge designs included development of the orthodox steam locomotive; steam-electric traction; gas-turbine-electric traction; high-speed steam motors, and various internal-combustion engines with electric, hydraulic, and mechanical transmission. The most favored design, from Lubeck Technical University, was made up of two 4-8-4 five-cylinder compound-expansion locomotives, placed back-to-back with a condensing tender between. The general disposition owed something to Kuhn's earlier design and the design of each locomotive was derived from Chapelon's Sorbonne proposal. None of the designs was realized, though one of them was based on the General Electric "Steamotive" which was tested in the U.S. in the late 1930s. Even more gigantic locomotives were designed for the broad-gauge project.

In 1939, the railway speed record was 230.2 km/h (143 mph) set between Ludwigslust and Wittenberge, Germany, in June 1931 by the propeller-driven Kruckenberg Rail Zeppelin. The record stood until February 1954. The Kruckenberg Car was an experimental vehicle to test vehicle behavior at high speeds and had a steerable axle controlled from the cab. In the 1930s the speed record for steam traction was 161 km/h set in November 1934 by 4-6-2 engine 4472 on the LNER between Grantham and Peterborough, England. This was raised in March 1935 to 173.8 km/h over the same route. A speed of 181 km/h was equaled by Milwaukee Road 4-4-2 and LNER 4-6-2 in 1935, and LMS 4-6-2 in 1937. In May 1936 German 4-6-4 O5.002 reached 200.4 km/h at Neustadt an der Dosse, and in July 1938 the record for steam traction was set at 202.8 km/h on the LNER between Grantham and Peterborough by an A4 4-6-2 engine. These records were less than the speed gained by electric traction in 1903 (Marienfelde-Zossen). They were surpassed by the diesel records of 181 km/h set by the Pioneer Zephyr in the U.S. in May 1934; and 205 km/h, reached by the Leipzig railcar between Ludwigslust and Wittenberge in February 1936. In June 1939, a Kruckenberg diesel set reached 215 km/h between

Hamburg and Berlin. Very-high-speed running then ceased during the war and in the long recovery period afterward. When research into high-speed railway operations was resumed in 1954, the initiative was largely with electric traction.

By the mid-1950s, the industrial economies had recovered sufficiently from the war to make use of airlines and motor transport on an increased scale. Railroads needed to increase speed of services to win back passengers from airlines and automobiles for journeys between 400 and 600 kilometers. It was necessary to determine how fast trains could be operated over existing routes within the standard loading gauge. A separate issue concerned the forms guided land transport might take. There were three general strategies: to run trains over existing railways at much higher speeds; to build new high-speed railways of conventional form, reserved for trains to run at very high speeds unhindered by slower operations; and to seek a completely new form of guided high-speed transport. The former strategy was pursued by British Rail via its Advanced Passenger Train project. This was a new, tilting electric train designed to take curves at higher than normal speeds, and to run at a maximum well above conventional train speeds; that is, faster than the high-speed diesel train introduced in the 1970s. The same concept has been used in the Spanish "Talgo" trains since the 1950s. It enabled the existing infrastructure to be used, but created difficulties in routing trains at lower speeds. Gas-turbine power was used during the experimental stage, but electric traction was chosen for production units. Unfortunately the project failed for nontechnical reasons. The strategy of building segregated high-speed routes (*Lignes a Grande Vitesse*, or LGV) was pioneered by the French and Japanese and taken up by other nations. Standard gauge was the norm, so these lines could be linked into the existing network in most instances, and the *Trains á Grande Vitesse* (TGV) could run over old and new lines and use city stations. On segregated routes, trains ran through curves at a set speed, so the rails could be superelevated to eliminate passenger discomfort. Tilting coaches were unnecessary and design was simplified. Segregated lines were the best means of getting high-speed rail services, but were extremely expensive and required state aid. Availability of funds sets the limits to the extent of the network. The third strategy, to build completely new kinds of guided land transport systems, has not been successful despite experiments with many different forms, none of which (with the possible exception of maglev, or magnetic levitation) approached the performance of the conventional TGV. Their proposers failed to anticipate the services offered on the electric LGV. The British guided or tracked hovercraft, the French jetrain, and the present-day German and French "maglev" projects are the best-known examples. At the close of the twentieth century, only maglev was being pursued with serious commitment and full-scale test facilities in Japan.

Research into very high-speed running was resumed in 1954. In February the SNCF ran locomotive hauled carriages on the 1500-volt DC Paris–Lyons main line. Locomotive CC 7121 reached 243 km/h with a three-coach train between Dijon and Beaune, thereby breaking all previous speed records. The rolling stock was relatively unmodified. In 1955, further SNCF speed trials were held on the electrified Bordeaux–Hendaye main line, with modified rolling stock and supply system. Mobile substations increased supply voltage from 1500 to 1900 volt DC. In March 1955, locomotive CC7107 reached 326 km/h, and BB9004 set a new record of 331 km/h, which lasted until 1981 when the SNCF TGV reached 371 km/h. The tests showed that a great deal had to be done to improve pantograph–catenary and vehicle–rail interactions, and to reduce aerial disturbance, which raised dust, disturbed ballast, and generated noise. Research was begun in several countries to achieve these goals and operate wheel-on-rail systems at increased speeds. The Japanese *Shinkansen* system originated in plans for a new electric railway between Tokyo and Osaka to provide greatly increased passenger-carrying capacity when the existing narrow-gauge link could no longer cope. A high-speed segregated line was opened in 1964 to provide a start-to-stop average of 161 km/h. Plans to work freight trains by night were abandoned: short nighttime periods were used for repairs; the rest was needed for passenger working. Success of the first line led to construction of the *Shinkansen* network, which stimulated similar high-speed railways elsewhere. The Japanese used lightweight sets with motors distributed throughout the train. Though successful, these networks were opposed on economic and environmental grounds. They required expensive noise suppression screens and embankments. State subsidies were essential. French trials in 1954–1955 resulted in the first regular passenger services in Europe to run at 200 km/h begun in 1967 by the *Capitole* express between Paris and Toulouse. These first trains used locomotive haulage and ran over ordinary railways with improved signaling, control, and communications. In the 1970s

these SNCF services were faster than any outside Japan. The need to reduce axle loading called into question the use of locomotives on LGV, but both locomotives and lightweight sets have found use. To reduce axle loading, and to provide very fast services on nonelectrified routes, the SNCF tested gas turbine-powered sets with hydromechanical transmission. In 1969, a five-car set was built with three trailers between two outer gas turbine-powered cars, containing traction alternators. The train was articulated, with traction motors on each axle. Eventually power-to-weight ratio was raised to 23.3 kilowatts per ton. Tests began in April 1972 and in December a speed of 318 km/h was reached. During four years running, this unit reached 250 km/h over a thousand times. Features now common on high-speed rolling stock were developed during these trials. The French plans to use such trains over existing railways met with routing problems, and eventually a network of segregated LGV was constructed. Changes in relative fuel costs prompted a switch to electric traction, using the new international standard of 25 kilovolts and 50 hertz. Some TGVs could also run under the 1500-volt DC of older lines. The French LGV, like the Japanese, employed cab signaling, and there were no lineside signals apart from marker boards at the start of each block. Signal blocks were indicated for five blocks ahead.

In Great Britain a different philosophy was pursued because new, segregated lines were ruled out on grounds of excessive costs and disruption to built-up areas. The High Speed Train (HST) was developed in the late 1960s to provide services at 200 km/h over improved existing lines using two diesel electric power units at the end of a rake of coaches. This proved most successful. To provide much higher speeds, of 250 km/h and above, the Advanced Passenger Train was developed, drawing on research into vehicle–rail interaction begun at the Railway Technical Centre, Derby, in 1964 to resolve problems related to high-speed goods operations. Such research showed that wheel-on-rail systems could operate at at least 320 or even 400 km/h. The present record is 515 km/h held by a French TGV. Because of this, British Railways Board rejected projects for unusual schemes and launched the Advanced Passenger Train (APT) program in 1968. The project received very little funding compared to that allocated by the French and Japanese states to the LGV and *Shinkansen*.

The engineers' story is related by Williams (1985). Special research and development facilities were set up in Derby and a disused line converted into a test route. Members were recruited to the design team from outside the railway industry. The APT experimental unit was completed in 1972 and used to test the tilt mechanism, bogies, suspensions, and train–track interaction. Authority for three prototypes was given in 1974. An electric and gas turbine-powered version were considered. However, the final test run was made in April 1976 and the project was terminated. Succeeding operations were left to the HST, and a new generation of conventional electric locomotives able to run at 225 km/h. Tilting trains have been used in Scandinavia, Italy, Switzerland, Norway, and Japan to reduce journey time over routes with heavy curves, but most TGVs at work in many lands are nontilting. The success of the Japanese and French TGV led to similar trains, working over routes partially or largely segregated, in Germany, Italy, Belgium, Netherlands, Spain, and South Korea. The cross-Channel Eurostar trains are an example. The TGVs associated with France and Germany have been used to advance the international business interests of the railway engineering industries of these countries, though globalization is reducing such national identification.

In the U.S., the primacy given to air transport for moderate to long distances and the reliance on automobiles for much else discourages construction of LGV on the European scale. In the closing years of the twentieth century, the U.S. introduced TGV, based on European practice, for service in the Northeast Corridor (Boston, New York, and Washington D.C.). These were tilting trains and could operate at 240 km/h. The future of LGV is uncertain, granted their great expense and the success of airlines in meeting competition over distances between 200 and 600 kilometers. The TGV work best between large centers of population where total journey times are about 3 or 4 hours maximum, and time is important.

See also **Railway Mechanics**

MICHAEL C. DUFFY

Further Reading

Anonymous. Les tres grandes vitesses ferroviaires en France. *Revue d'Histoire des Chemins de Fer*, 12–13, Paris, 1995. Special issue devoted to French TGV.

Hughes, M. *Rail 300*. Ian Allan. 1988.

Rahn, T., Ed. *ICE: High-Tech on Rails*. Hestra-Verlag, Darmstadt, 1986.

Tassin, Y., Nouvion, F., and Woimant, J. "*Histoire de la traction electrique, tome 2: De 1940 a nos jours*. La Vie du Rail, Paris, 1986.

Williams, H. *APT: A Promise Unfulfilled*. Ian Allan, 1985.

Rail, Steam Locomotives

In 1900 British influence on the design of steam locomotives was strong but existed alongside distinct French, German, and American traditions. In the twentieth century, American designs and organization of steam traction became global standards due to research by universities, railroads, and the engineering industry. After 1900, North American designs departed radically from their European counterparts and became heavier in size, weight, tractive effort, and power. The basic "Stephenson" locomotive remained the norm throughout the twentieth century and was improved by superheating, piston valves, and the rational proportions pioneered by W.F.M. Goss and William Woodard in the U.S. Radical departures from this norm failed and were always rare. For general-purpose railways outside rapid transit systems, there was no alternative to steam traction before the 1940s. Throughout the century, express passenger locomotives worked heavier trains at sustained speeds higher than in the period from 1890 to 1910, but top speeds did not increase much apart from a few record runs of over 160 kilometers per hour (km/h), and steam traction coped until the 1950s. (The world speed record for steam was set at 203 km/h in 1935 by the British 4-6-2 engine Mallard.) From 1900 to 1930 freight trains of 3000 to 6000 tons were moved at 8–16 km/h mph and required engines with high tractive effort, low speed, and moderate power. The compound-expansion Mallet type, weighing up to 350 tons, was employed in the U.S. for this duty. After 1930, operations favored lighter, faster trains, and the high-powered, high-speed, rigid-framed 2-8-4 or 2-10-4 wheel arrangement type was widely used.

Woodard, Goss, and others increased performance using the basic Stephenson type, demonstrated by engine No. 50,000 of 1910. In Britain, the locomotives designed by H.N. Gresley after 1922 were derivatives of engine 50,000. Woodard's philosophy was applied to 4-6-2, 2-8-4, 2-10-4 and the later Mallet types, and by 1930 the American steam locomotive had virtually reached its final form. Locomotives exported to Russia, Japan, and China established the American exemplar in those countries in the 1930s. High-powered designs reached finality in the S1 6-4-4-6 Duplex locomotive of 1939, built to sustain 6500 horsepower and speeds of 160 km/h mph with 1000 tons on the level. Operated on the Pennsylvania Railroad, its great size (43 meters long), weight (482 U.S. tons), and limited route availability suggested that conventional steam traction was approaching its limit and that another mode of traction would be needed to raise power and speed standards after the 1940s.

Steam traction reached practical perfection by integrating the locomotive with the track through better balancing technique; rationalizing locomotive proportions, and applying scientific management to all parts of the traction system. This system worked best with simple, standardized locomotives with two outside cylinders and all components distributed for ease of maintenance. Mixed traffic locomotives were favored; that is, those able to work both passenger and freight trains. This was essential in the 1930s Depression years, during the 1939–1945 World War II years, and in the postwar period when labor was costly and funding scarce.

Locomotive science advanced after 1930 and led to new forms of steam locomotive using innovations in power station technology and marine practice. These combined high-pressure boilers with turbines, vacuum condensers, and electric transmission. Extremely complex machines were tried without success. Many projects were delayed by the war and revived afterward, when diesel and electric traction were proven and affordable. In France, Chapelon developed a high-performance locomotive, based closely on the orthodox form, which used compound expansion. Remarkable performances were demonstrated by Chapelon's 4-8-0, 4-6-2, 4-8-4, and 2-12-0 types between 1930 and the late 1940s. These were more promising designs than the contemporary high-pressure units developed by Muhlfeld in the U.S., but they had no significance once postwar economies permitted use of diesel and electric traction.

In the 1940s and early 1950s final efforts were made to develop a new form of steam locomotive free from the defects intrinsic to the orthodox type. The steam-electric locomotives of the Chesapeake and Ohio, and Northern and Western railways, tried between the late 1940s and early 1950s, were the last efforts made in the U.S. The Northern and Western railway retained steam traction using well-planned servicing stations, but the system became anachronistic, and belated dieselization proved essential in the 1960s. In the British Isles the unsuccessful experiments of Bulleid on locomotives with steam motors mounted in bogies concluded the search for a new form by the mid-1950s. Steam traction ended with the acceptance of the North American model, worked on the common-user system according to principles of scientific management. Monthly "best mileages" increased from about 6,000 miles (9656 kilometers) in 1910 to 12,000 miles (19,300 kilometers) per month in 1939

for 4-6-4 locomotives on the high-speed Chicago and Northwestern's Hiawatha services, and 24,000 miles (38,600 kilometers) per month for 4-8-4 locomotives on New York Central express trains in the late 1940s. The need for increased speeds and productivity in an age of motor transport, cheap air services, and increasing labor costs accelerated the shift to diesel and electric traction. Steam traction was intrinsically obsolete, and all attempts to save it, including one-man operations, automatic boiler regulation, multiple-unit working, and electric transmission, failed. The problem remained that the old forms were destructive of the track at high speeds, as the Chicago and Northwestern rail working tests had shown as early as the 1930s.

In its final form the American exemplar set the world standard. Typical weights for a North American 4-8-4 in running order were:

Engine 220,000–230,000 kilograms
Tender: 160,000 to 210,000 kilograms
Length: 33 to 37 meters

Sustained cylinder power was 5500 to 6000 horsepower, with a maximum of around 7000 horsepower. The largest articulated locomotives (e.g., Union Pacific 4-8-8-4 class) had an engine weight of 327,600 kilograms; tender of 198,000 kg; and a length of 40 meters. A modern steam locomotive would have most of the following features: a cast-steel bed; a welded boiler and steel firebox; rocking grate; self-cleaning smokebox; regulator valves in superheater header; compensated springing; mechanical lubrication; antidistortion disk wheels; roller bearings on all axles; roller bearings in rods and motion; and modern front end. Some engineers would add a combustion chamber, syphons and circulators, and Franklin wedges in the driving axle boxes. Most American engines would have a stoker (unless they were oil-burning), power reverse gear, and two outside cylinders. The final types built later in Europe, Russia, and China were derived from this American exemplar of the 1940s, but were much smaller.

Improvements to the steam locomotive were applied in backward economies with cheap labor until the 1990s. The work of David Wardale in South Africa and Livio Dante Porta in South America provide examples. Other improvements, like the Giesel ejector invented in the late 1940s in Austria, enabled older engines to run more economically in the years left before scrapping. Recent improvements incorporated into new steam locomotives built in the 1990s for tourist railways do not change the status of steam traction. The fuel crises of the 1970s and later led to plans for coal-fired locomotives, which included triple-unit steam-electric machines with condensing tenders, but none was constructed, and the coal-fired diesel electric and the coal-fired gas-turbine are favored instead. The schemes for atomic-powered steam-electric locomotives, proposed in the late 1940s and 1950s were never taken seriously by railwaymen.

See also **Rail, Diesel and Diesel Electric Locomotives; Rail, Electric Locomotives; Rail, High-Speed; Railway Mechanics**

MICHAEL C. DUFFY

Further Reading

Atkins, P. *Dropping the Fire: The Decline and Fall of the Steam Locomotive*. Irwell Press, 1999.

Ahrons, E.L. *The British Steam Railway Locomotive 1825–1925*. Locomotive Publishing, London, 1927. Reprinted by Ian Allan, 1966.

Carling, D.R. Locomotive testing stations. *Trans. Newcomen Soc*., 42, 105–182, 1972.

Dalby, W.A. *The Balancing of Locomotives*. Edward Arnold, London, 1930.

Duffy, M.C. Rail stresses, impact loading & steam locomotive design, in *History of Technology*, 9th ann. vol. Mansell, London, pp. 43–101.

Durrant, A.E. *The Mallet Locomotive*. David & Charles, Newton Abbot, 1974.

Gresley, H.N. High pressure steam locomotives. *Proc. Inst. Mech. Eng*., 121, pp. 101–206, 1931.

Hodgson, J.T. and Lake, C.S. *Locomotive Management: Cleaning, Driving, Maintenance*. Revised by Oaten, W.R., Tothill Press, London, 1954. Detailed, illustrated description of modern British locomotives. A standard railwayman's text.

Johnson, R.P. The Steam Locomotive: Its Theory, Operation & Economics including comparisons with Diesel-Electric Locomotives. Simmons-Boardman, New York, 1944, reprinted 1981.

Lucas, W.A., Ed. *100 years of Steam Locomotives*. Simmons-Boardman, New York, 1957. Evolution of American locomotive shown through engineering drawings.

Nock, O.S. *The British Steam Railway Locomotive 1925–1965*. Ian Allan, 1969.

Ransome-Wallis, P. *Concise Encyclopedia of World Railway Locomotives*. Hutchinson, 1959. Contains survey of final generation of steam locomotives.

Sauvage, E and Chapelon, A. *La Machine Locomotive*. Beranger, 1947. Reprinted by Les Editions du Layet, 1979.

Wiener, L. *Articulated Locomotives*. Constable, London, 1930.

Railway Mechanics

Railway mechanics grew from the nineteenth-century analysis of the union of locomotive and track. Improving the "fit" between engine and track was essential for safe and progressive opera-

tion. Analysis of the new machine stimulated general engineering mechanics. Investigators studied the balancing of locomotive mechanisms in the nineteenth century, and the phenomenon of fatigue failure was first identified in a locomotive context. In the early twentieth century, the study of shatter cracks in rail heads resulted in new regulations governing the manufacture of rails. Sylvester and Tschebyshev considered the kinematics of mechanisms, including valve gears. O. Reynolds studied balancing of reciprocating parts, and coupling rod failure. Railway studies of materials are reflected in the work of Mohr, whose analysis is still used in strength of materials.

In the 1890s, mechanics was supplemented by thermodynamics, and quantified analysis of energy flows were carried out in locomotives. The electric power industry used energy analysis, which passed into railway engineering after 1900. In the 1880s A. M. Wellington published his general theory of railway location, linking engineering and operation to climate, topography, commerce and economics. The influence of this work is still felt. In 1900, the machine-ensemble (traction system) was integrated closely with signaling via interlocking. Systematic investigations into locomotive proportions were conducted by W.F.M. Goss at Purdue, following earlier studies by Borodin in Russia, and Le Chatelier in France. Le Chatelier was a pioneer of rational management, efficient energy use, and scientific organization, as were Taylor, Gilbreth, Gantt, and Ford. Their theories transformed railway engineering management and design, and American exemplars transformed every department of the machine ensemble: traction, permanent way, signaling, and control. New devices were in part responsible: track design, rail shape, track circuits, power signaling, automatic signals, mechanized equipment in depots, and electrical safety mechanisms; but revolutionary concepts in management and organization drove much of the technical changes. Electric railways witnessed early attempts to model the technical and economic features of large industrial systems.

The work of Wellington in the U.S. modeled the railway in its general environment, and the work of G. W. Lomonossoff in Russia resulted in a comprehensive theory of railway mechanics in the early twentieth century. He advocated research departments dedicated to railway mechanics to frame rational strategies for engineering development. Lomonossoff's theories and experiments showed the difficulty of assigning accurate, measured values to parameters in theoretical expressions, without which comprehensive theories were of little value. Modern testing techniques were developed to overcome this problem. Experiment and experience were insufficient to solve major problems, and theoretical analysis was needed. The Russian contribution was considerable. Grinevetsky, Syromyatnikov, Nikolayev, and Shelest developed academic courses on locomotive design between 1900 and the 1960s. Similar academic work was done in Germany and the US. Steam locomotive science was developed in France by Chapelon and by William Woodard and R.P. Johnson in the U.S.

The effects on track of out-of-balance forces in steam locomotives was made in the 1940s tests carried out on the Chicago and Northwestern Railroad. The behavior of rails, trackbed, and locomotive components were measured using electrical recording apparatus and high-speed cameras. Wheel lift was observed during slipping. A mechanical model of wheel–rail interaction was constructed using rollers and springs, which could represent different rail dimensions, ballast characteristics, axle loads and train speeds. This helped compare theory with observations. Guidelines were derived governing springing, wheel design, balancing and number of cylinders. The tests were repeated under British conditions by the LMS railway in 1941.The LMS set up a research center in the 1930s after the American model, which investigated train resistance, aerodynamics, and vehicle–rail interaction.

F.W. Carter pioneered general analysis of rail–vehicle interaction in 1916. Hunting or sustained oscillations due to dynamic instability lacked adequate theoretical interpretation, as did creep and the dynamics of curving; these affected all vehicles on all kinds of track. The growth of oscillations at high speeds, even on straight track, limited very high-speed operations. Graphical analysis was often needed to solve equations. Carter's work, though of great importance, was limited in value because the configurations of vehicle body, wheels and trucks that he considered were inherently unstable. The analysis was subsequently developed by Rocard in France. Many railway engineers tolerated hunting as inevitable, but the high-speed tests of the 1930s in Germany, Great Britain, and the U.S. drew attention to the lack of understanding. In Great Britain the LMS railway contracted Prof. Inglis at Cambridge University. The Cambridge work, published in 1939, included equations of motion for a two-axle bogie, but the twelfth-order expressions could not be easily solved by the techniques then available. Advances were made after World War

II by Matsudaira at the Railway Technical Institute, Tokyo, who developed the concept of stiffness to stabilize spring-restrained wheelsets at any critical speed. By the 1960s this work enabled the Japanese railways to operate very high-speed *Shinkansen* trains. The problems investigated concerned all forms of train, and in the 1960s British Railways contracted Imperial College to develop adequate analysis. In both the Japanese and British work, dynamicists with a background in aeroelasticity played vital roles (e.g., Matsudaira and Wickens). The theoretical analysis was tested by experiment using roller test stands and instrumented trains and track, greatly aided by the availability of computers to solve expressions that were difficult or impossible to solve by earlier methods. Developing adequate theories of vehicle behavior on curves was difficult and had been investigated from 1883 to 1903 in Britain, France, and Germany. A fundamental contribution was made in Britain by Porter whose 1934 studies on mechanics of a locomotive on curved track became classics. By the late 1960s, fairly complete theories of curving were in existence, which included vehicle form, creep, and stability. In 1977, a general theory of curving was formulated by Elkins and Gostling.

At the end of the twentieth century, research and development involved cooperation among states, manufacturers, universities, and railway companies. The trend was toward international cooperation, encouraged by globalization of leading manufacturers. The German program in advanced railway technology, begun in 1972, is one example. Government funding varied between 50 and 100 percent depending on the project. Federal funding in the U.S. also varied with the project. The German program covered vehicle and rail interaction, vehicle design (including aerodynamics and control), track and trackbed design and behavior, control technology, power technology including pantograph-contact wire interaction, and environmental matters. The ME-DYNA program was structured for nonlinear simulation of vehicle track interaction. The Munchen–Friedmann test facility was constructed from 1977 to 1980 to investigate four-axle vehicles. Wheelsets have been tested up to 503 km/h; techniques for predictive analysis of bodywork stresses, modes of vibration, and distortion during all modes of motion were all improved; and aerodynamic behavior of vehicles and pantographs was optimized. Without this analysis, dependent on electronics for its execution, the modern railway, with its enhanced performance and close integration of traction system, signaling, and automatic control would not be possible.

See also **Rail, Diesel and Diesel Electric Locomotives; Rail, High Speed; Rail, Steam Locomotives**

MICHAEL C. DUFFY

Further Reading

Carter, F.W. *Railway Electric Traction*. Arnold, 1922.

Carter, F.W. On the stability of running of locomotives. *Proc. Royal Soc. A.*, 121, 585–611, 1928.

Duffy, M.C. Rail stresses, impact loading, and steam locomotive design, in *History of Technology*, vol. 9. Mansell, London, 1984, pp. 43–101.

Gilchrist, A.O. The long road to solution of the railway hunting and curving problems. *Proc. Inst. Mech. Eng.* 212, Part F, 219–226, 1998.

Lomonossoff, G.V. *Introduction to Railway Mechanics*. Milford/Oxford University Press, Oxford, 1933.

Majumdar, J. Economics of Railway Traction. Gower, 1985.

Rahn, T., Hochbruck, H., Moller, F.W. and Lubke, D., Eds. *ICE: high-tech on rails*. Hestra-Verlag, Darmstadt, 1986.

Wellington, A.M. *The Economic Theory of Railway Location*. Wiley/Chapman & Hall, 1887, 1889.

Wickens, A.H. The dynamics of railway vehicles—from Stephenson to Carter. *Proc. Inst. Mech. Eng.*, 212, Part F, 209–217, 1998.

Rectifiers

Rectifiers are electronic devices that are used to control the flow of current. They do this by having conducting and nonconducting states that depend on the polarity of the applied voltage. A major function in electronics is the conversion from alternating current (AC) to direct current (DC) where the output is only one-half (either positive or negative) of the input. Rectifiers that are currently, or have been, in use include: point-contact diodes, plate rectifiers, thermionic diodes, and semiconductor diodes. There are various ways in which rectifiers may be classified in terms of the signals they encounter; this contribution will consider two extremes—high frequency and heavy current—that make significantly different demands on device design.

Diodes assumed a special importance after the invention of wireless. The conventional method of transmitting audio signals was to use amplitude modulation (developed by Reginald A. Fessenden in 1906). The wireless signal consisted of a single high-frequency signal (carrier wave), whose amplitude was changed in direct proportion to the audio signal that was to be transmitted. The carrier frequency is well outside the range of the human ear. However, once the signal has been passed through a detector, such as the silicon point

contact diode patented by Greenleaf Whittier Pickard in 1906, the much slower changes of the positive "envelope" can be heard with the aid of headphones.

In the days before transistors, the amplification of the acoustic signal was achieved using thermionic valves such as the triode invented by Lee de Forest in 1906. These devices required a high-voltage DC supply. This was normally derived from the 50- or 60-hertz AC mains via one or more thermionic diodes (invented by John Ambrose Fleming in 1904). A metal plate and an electric filament are separated within a vacuum enclosure. When a current is passed through the filament (normally tungsten), electrons are emitted from its surface. If the voltage on the metal plate (called the anode) is positive, then current will flow. However, if the voltage on the plate is negative, then electrons are repelled and current does not flow.

Researchers L.O. Grondahl and P.H. Geiger discovered the copper oxide rectifier in 1927 and noted that the flow of current through a layer of copper oxide on metallic copper is not symmetrical with respect to bias; more current flows if a battery is connected one way than if it is connected with the reverse polarity. Similar observations were made about nickel oxide in contact with nickel, but the market for low-power, low-frequency applications was dominated by selenium rectifiers, which were invented by Charles E. Fitts in 1933. Selenium has a reverse breakdown voltage in the region of 20 to 30 volts, so higher voltages such as the AC mains supply could be converted to DC using a series of selenium diodes separated by metal plates. A stack of such diodes could be arranged as a bridge for full-wave rectification and as such were used in radio and TV receivers into the1960s.

The advent of silicon *p–n* junction diodes and transistors marked a revolution and consigned thermionic diodes and other valves to the history books. However, variants of some of the old devices are still in use today. The cat's whisker was a rudimentary type of metal-semiconductor diode. Structures of this type were extensively studied by the German physicist, Walter Schottky. Unlike ordinary semiconductor diodes the electrical conduction in this diode involved one carrier type only and this allowed them to switch very quickly between conducting and nonconducting states. It was therefore possible to use them at much higher frequencies, well outside the range of conventional junction diodes. Suitably designed point contact diodes are still used in radar and microwave telephone links. A particularly fast device called the "snap-off" diode is an essential component in frequency multiplier circuits and at one time this was the only way to generate signals in the range 50 to 90 gigahertz.

The conversion of mains supply to DC for heavy current applications presented a particular problem for rectification. The arc lamps in cinema projectors were one such example and the mercury-arc rectifier valves that were used were enormous in terms of volume occupied per amp supplied. Silicon junction power diodes were a relatively late development within the semiconductor revolution. Large currents required large-area silicon dice. In the early stages, the level imperfections, measured as defects per unit area in the starting crystal was sufficiently large that the yield of good-quality marketable devices was unacceptably low. Technological developments changed this situation and it is now possible to obtain diodes that can block in excess of 2000 volts when reverse biased. Low- and medium-power devices with slow or fast turn-off are now readily available at reasonable prices.

A silicon *p–n* junction diode has a property called the forward voltage drop. This is generally between 0.6 and 0.7 volts and would mean that if a diode were handling 1000 amperes, one would need to dispose of heat equivalent to approximately 700 watts, and this continues to present a challenge for design engineers. However, a variant of the original metal-semiconductor junction, called the Schottky diode, has two very useful properties. The forward voltage drop is significantly less than for *p–n* junction diodes, so that the equivalent size of diode package can carry much larger currents. Second, because they have very fast turn-off times, they can rectify high-frequency signals. Accordingly, they are a key component in the switched-mode power supply unit (SMPSU), a revolution in rectification. The incoming main is rectified using a conventional bridge circuit. This is then converted back to AC, but at very high frequency. High-frequency transformers are much lighter and more efficient than conventional 50- or 60-hertz transformers. The output from such a transformer is rectified using a Schottky diode. The power supply in a desktop computer will have several transformer or Schottky pairs, and a typical 250-watt PSU may have to deliver up to 32 amperes at 5 volts, 0.5 amperes at -5 volts, 10 amperes at 12 volts, and 0.5 amperes at -12 volts.

The thyristor is the ultimate in rectification. It has a gate contact that can be operated by currents between 10 microamperes and 200 milliamperes. The device remains "off" if the gate has not been triggered. The application of a current pulse to the gate during a positive half-cycle of the applied AC

voltage "fires" the thyristor. Current then flows between the anode and cathode terminals from that instant until the end of the half-cycle. This provides a variable level DC output. The thyristor is used in domestic lamp dimmers. Stacks of thyristors, each rated in excess of 1000 amperes, with a 2000-volt reverse breakdown are used in the high-voltage DC link between Britain and France.

See also **Control Technology, Electronic Signals; Radio Receivers, Crystal Detectors and Receivers; Semiconductors, Compound; Transistors**

DONARD DE COGAN

Further Reading

Braun, E. and Macdonald, S. *Revolution in Miniature: The History and Impact of Semiconductor Electronics*. Cambridge University Press, Cambridge, 1982.

Grondahl, L. O., and Geiger, P. H. A new electronic rectifier. *Proceedings of the American Institution of Electrical Engineers, Winter Convention, New York, 1927*, p. 357.

Refrigeration, Absorption

William Cullen's seminal experiments on the evaporation of liquids in the 1750s served as a precursor to the development of absorption refrigeration. Cullen, who accepted a chair in chemistry at the University of Edinburgh in 1756, demonstrated the possibility of refrigeration when he produced a small amount of ice through vaporization by reducing the pressure under a bell jar. One of Cullen's students, Joseph Black, made an important contribution to both physics and refrigeration engineering with his articulation in 1761 of the theory of latent heat, which accounts for the release and absorption of heat as a substance changes state. Another of Cullen's students at Edinburgh, Edward Nairne, improved upon the pioneer's laboratory experiments by placing a cup of sulfuric acid along with a cup of water under a partially evacuated bell jar. The sulfuric acid absorbed the vaporized water and thereby hastened the evaporation process. Nairne's successor at Edinburgh, John Leslie, and later a British brewer named John Vallance made further improvements to Nairne's sulfuric acid–water absorption device in the early part of the eighteenth century, but neither developed their devices beyond the experimental stage.

Edmund Carré, a French inventor, developed a sulfuric acid absorption refrigeration device in the 1850s. Although it was installed in a number of bars and restaurants in France, the device malfunctioned and underwent disabling corrosion. Edmund's brother, Ferdinand Philippe Eduard Carré, concluded that the problem was not the device itself, but the choice of refrigerant. By replacing the sulfuric acid and water mixture with an ammonia and water mixture, Ferdinand Carré developed the first successful absorption refrigeration system. The initial Carré device had just two components. One component combined the functions of a generator and an absorber; the other combined the functions of a condenser and an evaporator. Because of the dual functions of each component, the machine operated intermittently rather than continuously. Soon thereafter, however, Carré developed a continuous machine.

The Carré absorption unit had all of the components of a modern day absorption machine, including: an evaporator, an absorber, a generator, an expansion valve, and a pump. In ammonia–water absorption systems, the ammonia refrigerant flows at low pressure through the evaporator where it absorbs heat from the refrigerator cabinet. It then moves to the absorber where it mixes with water. The water and ammonia mixture is then pumped into the generator, and the heat of the generator separates the mixture into water and high-pressure ammonia gas. The water returns to the absorber, and the ammonia gas then flows into the condenser, where cool water passing over the outside of the coils removes heat from the ammonia and allows it to return to a liquid form. From there, the ammonia gas moves through an expansion valve and is ready to flow back into the evaporator where, once again, it absorbs heat from the refrigerator cabinet (see Figure 25).

The absorption machine gained quickly in popularity. It was adopted and produced in Germany, Great Britain, and the U.S. Carré refrigeration units were shipped through Union blockades during the U.S. Civil War to a military hospital in Augusta, Georgia, and King Ranch near Brownsville, Texas. Manufacturers readily improved upon the Carré design. The use of wood chips to heat the generator was the most glaring problem with the machine. To gain greater control over the heating process, manufacturers of industrial machines replaced the wood-burning system with steam coils. Manufacturers developed more efficient rectifiers and new types of absorbers throughout the late nineteenth and early twentieth centuries. The refrigeration units were developed almost exclusively for large-scale commercial establishments, such as ice-making plants and breweries. The production process, therefore, resembled a construction operation. Without the

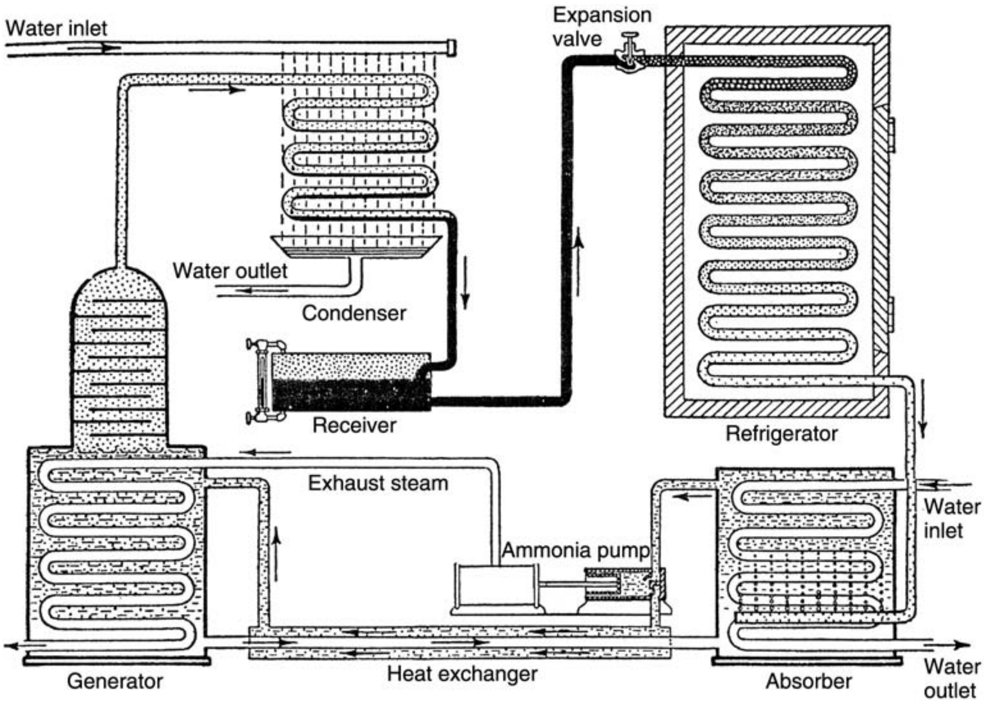


Figure 25. Absorption refrigeration machine.
[*Source: From Anderson, O.E. Refrigeration in America: A History of a New Technology and Its Impact. Princeton University Press, Princeton, 1953.*]

possibility of stable designs, refrigeration companies depended on specialty manufacture, craft skill, and on-site engineering improvements to build their systems.

A number of characteristics distinguish absorption refrigerators from mechanical-compression systems. Both force the evaporation and condensation of a refrigerant to transfer heat across physical boundaries, but the absorption system uses heat to generate the pressure necessary to enable the evaporation process, while the mechanical-compression system relies on an electric-powered compressor. Rather than mechanical processes and electrical energy, the absorption system depends on heat and physiochemical processes for its operations. The absorption system, furthermore, has more components than its competitor, has no moving parts, and has a refrigerant consisting of two substances, not one.

Absorption refrigeration remained competitive with compression systems up to World War I in the U.S. During the interwar years, however, absorption refrigeration for both industrial and domestic settings plummeted in popularity. The higher initial cost of absorption machines and the inability of manufacturers to match the innovations of producers of compression systems are only part of the story. Household absorption refrigerators depended on gas heat, while the compression systems depended on electrical power. The emerging giants of the electrical industry in the U.S.—General Electric, Westinghouse, and the Frigidare Division of General Motors—and their well-financed partners in the electrical utility industry were able to muster far greater financial and technical resources than producers of absorption refrigeration systems. The single major producer of absorption refrigerators, Servel, was at too great a financial disadvantage to compete with the electrical industry in research and development, and marketing. Absorption refrigeration, however, did not vanish from the marketplace. Despite its early eclipse, it occupied an important niche in the global refrigeration market throughout the twentieth century.

See also **Air Conditioning; Refrigeration, Mechanical**

GLEN ASNER

Further Reading

Anderson, O.E. *Refrigeration in America: A History of a New Technology and Its Impact*. Princeton University Press, Princeton, 1953.

Cowan, R.S. *More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave*. Basic Books, New York, 1983.
Sauer, H.J. Jr. Refrigeration, in *Encyclopedia of Energy, Technology, and the Environment*, vol. 4, Bisio, A. and Boots, S., Eds. Wiley, New York, 1995.
Thévenot, R. *A History of Refrigeration Throughout the World*. Translated by Fidler, J.C. International Institute of Refrigeration, Paris, 1979.
Woolrich, W.R. *The Men Who Created Cool: A History of Refrigeration*. Exposition Press, New York, 1967.

Refrigeration, Mechanical

Over 170 years elapsed between William Cullen's experiments on refrigeration at the University of Glasgow and the mass production of mechanical-compression refrigerators for the consumer market in the 1920s. Beginning with Cullen in 1748, scientists and inventors across the Western world manipulated the basic properties of the elements to produce ice and cold air. The most prominent methods of refrigeration included air-cycle compression, vapor compression (or mechanical compression), thermoelectric, and absorption. While no single method proved superior in all circumstances, mechanical-compression refrigeration emerged by the 1940s as the dominant method for both domestic and industrial refrigeration.

Inventors throughout the Western world contributed to the development of mechanical-compression refrigeration throughout the nineteenth century. Oliver Evans, the American inventor of the automated flour mill, laid the foundation for the continuous-cycle vapor-compression refrigerator (see Figure 26) when he conceived a method in 1805 for recycling vaporized refrigerant. After removing heat from the surrounding environment, vaporized refrigerant would move through a compressor, and then a condenser, where it would revert back into a liquid form and begin the process again. Although Evans failed to transform his idea into a practical device, Jacob Perkins, an American-born resident of London who befriended Evans during a stay in Philadelphia, built on Evans' ideas to construct a cyclic vapor-compression machine in 1834. The Perkins machine, which used ether as a refrigerant, was the first full-scale machine to contain a compressor, a condenser, an expansion valve, and an evaporator—the basic parts of the modern mechanical-compression refrigeration system. Perkins, however, never developed his machine for commercial use.

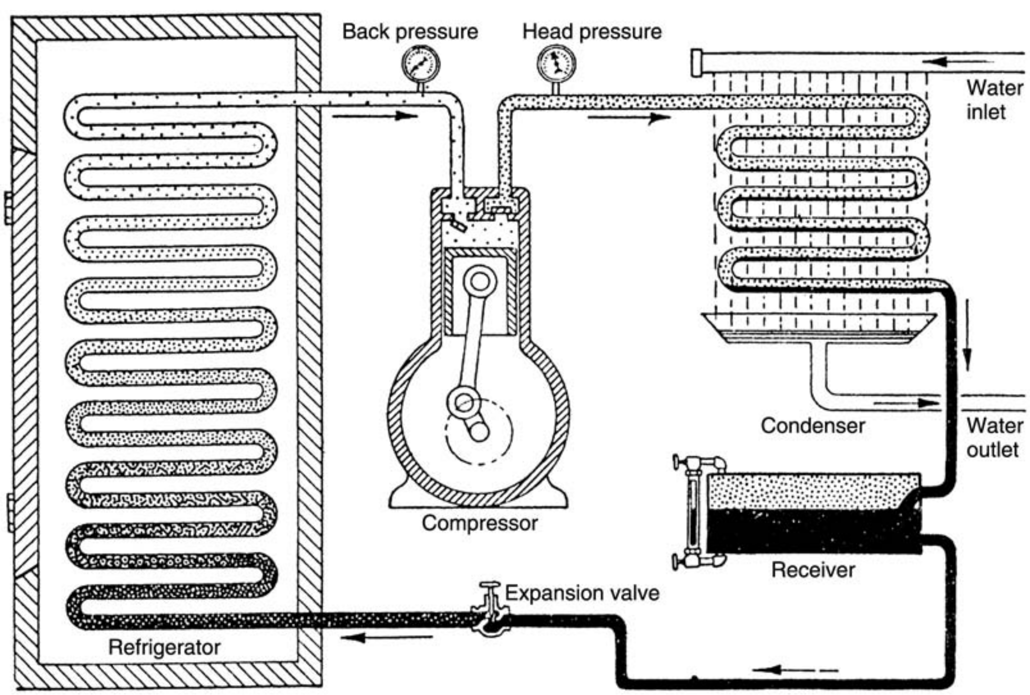


Figure 26. Vapor-compression refrigerator.
[*Source: From Anderson, O.E. Refrigeration in America: A History of a New Technology and Its Impact. Princeton University Press, Princeton, 1953.*]

Dr. John Gorrie, the director of the U.S. Marine Hospital in Apalachicola, Florida, received a patent in 1851 for the first refrigeration system operated for practical use. Gorrie, however, installed his machine only at his own hospital, and it was an air-cycle compression machine rather than a vapor-compression machine. Whereas the refrigerant in a vapor-compression machine alternates between a gaseous and a liquid state, the refrigerant in an air-cycle compression machine—air—remains gaseous throughout the compression and expansion cycles. Alexander Twining, a college professor and practicing civil engineer, received a patent in 1853 and built the first commercially viable continuous-cycle vapor-compression refrigeration machine based on Jacob Perkins' original design. Several inventors subsequently developed large-scale compression machines for industrial settings that varied from Twining's machine only in component design and choice of refrigerant. Many of the early machines used ether, a highly volatile and toxic substance, as a refrigerant. In the 1870s, David Boyle in the U.S and Carl von Linde in Germany introduced ammonia, a less toxic refrigerant that changed states more readily than ether. However, other more toxic refrigerants remained in use alongside ammonia. Reciprocating compressors were first developed in 1880 for industrial applications, such as ice making, brewing, and meat packing.

Demand for domestic refrigeration emerged by the 1880s, as indicated by the widespread use of iceboxes in households. Rudimentary mechanical refrigeration units for the home first appeared in 1910, and the Kelvinator Corporation began quantity production of the first domestic refrigerator unit with automatic controls in 1918. These early machines, however, were unreliable, expensive, plagued with technical problems, and even dangerous. Most systems used sulfur dioxide or methyl chloride as refrigerants, both of which are highly toxic. Even less-toxic ammonia could prove lethal if it leaked, and leakage was a constant problem with the early machines. They also required regular servicing due to chronic problems with thermostats, motors, and compressors. The separation of the refrigerating machinery from the refrigeration compartment exacerbated these problems by forcing the compressor to work harder.

The redesign of the refrigerating unit and the development of a new class of refrigerants helped to increase the reliability and safety of mechanical-compression refrigerators. General Electric developed the first hermetically sealed motor-compressor for domestic refrigeration in the mid-1920s. The GE Monitor Top, which was based on a 1905 design by Audiffren in France, contained all its mechanical parts in a single unit, ingeniously placed on top of the refrigerator box. The machine was air cooled and made of steel rather than wood, which was commonly used for most machines at the time. Other manufacturers soon replicated GE's design. By 1940, almost all domestic refrigerators were self-contained units and made of steel with no external parts.

The discovery of chlorofluorocarbon (CFC) refrigerants by Thomas Midgely and a team of researchers at General Motors in the late 1920s eliminated the immediate health and safety hazards of refrigeration. The company announced its discovery in 1930. Soon thereafter, the Kinetic Chemical Company, a joint venture of GM and DuPont, began manufacturing the CFC refrigerant known commercially as Freon. Although General Motors initially intended to restrict use of Freon to machines produced by its Frigidaire Division, the benefits and the potential profits from widely distributing the nontoxic and nonflammable refrigerant were too great to resist. Within a few years Freon became the refrigerant of choice for mechanical-compression refrigerators.

Other important landmarks in the history of refrigeration include the development of the humidity drawer for fruits and vegetable in 1930, General Electric's 1939 introduction of the dual temperature refrigerator, which had separate compartments for chilled and frozen foods, and Frigidaire's replacement of reciprocating compressors with quieter and smaller rotary compressors in 1933. Automatic defrosting was introduced in the early 1950s.

Whereas immediate health and safety concerns stimulated the development of chlorofluorocarbon refrigerants, long-term environmental concerns led to the abandonment of CFCs. In 1974 two chemists at the University of California at Irvine, F. Sherwood Roland and Mario Molina, discovered that chlorine atoms catalytically break down ozone when exposed to high-frequency ultraviolet light. Roland and Moline concluded that CFCs were stable enough to pass through the troposphere but would decompose and release chlorine in the stratosphere, where they would deplete the earth's ozone shield. Not until the British Antarctic Survey discovered a hole in the ozone layer in 1985, however, did concern over CFCs become widespread. In keeping with their commitments under the 1987 United Nation's Montreal Protocol, industrialized nations abandoned the use of CFCs by 1996. The search for refrigerants to

replace CFCs yielded new chemical compounds, such as hydroflourocarbons (HFCs), which were less damaging to the ozone layer. However, no single refrigerant emerged as a global standard. Research on refrigerants and refrigeration methods at the end of the twentieth century was as vibrant as it was at anytime during the century.

See also **Air Conditioning; Cryogenics; Refrigeration, Absorption; Refrigeration, Thermoelectricity**

GLEN ASNER

Further Reading

Anderson, O.E. *Refrigeration in America: A History of a New Technology and Its Impact*. Princeton University Press, Princeton, 1953.
Cooper, G. *Air-Conditioning America: Engineers and the Controlled Environment, 1900–1960*. Johns Hopkins University Press, Baltimore, 1998.
Cowan, R.S. *More Work for Mother: The Ironies of Household Technology From the Open Hearth to the Microwave*. Basic Books, New York, 1983.
Hård, M. *Machines are Frozen Spirit: The Scientification of Refrigeration and Brewing in the 19th Century—A Weberian Interpretation*. Campus Verlag, Frankfurt am Main; and Westview, Boulder, CO, 1994.
Thévenot, R. *A History of Refrigeration Throughout the World*. Translated by Fidler, J.C., International Institute of Refrigeration, Paris, 1979.
Sauer, H.J., Jr. Refrigeration, in *Encyclopedia of Energy, Technology, and the Environment*, vol. 4, Bisio, A. and Boots, S., Eds. Wiley, New York, 1995.
Woolrich, W.R. *The Men Who Created Cool: A History of Refrigeration*. Exposition Press, New York, 1967.
Woolrich, W.R. The history of refrigeration: 220 years of mechanical and chemical cold, 1748–1968. *ASHRAE Journal*, July, 31–39, 1969.

Refrigeration, Thermoelectricity

The scientific principles underlying thermoelectric refrigeration were understood by the mid-nineteenth century. In 1822, German scientist Thomas Johann Seebeck discovered that a needle would move when held near the junction of two dissimilar metals maintained at different temperatures. Seebeck misidentified the effect as magnetic, but Hans Christian Oersted, the father of electromagnetism, and James Cumming, a Cambridge chemist, correctly categorized Seebeck's discovery, known as the Seebeck effect, as an electrical phenomenon. A Parisian clockmaker, Jean Charles Athanese Peltier, made the second important discovery in the field of thermoelectricity in 1834 while performing an experiment to measure the conductivity of bismuth and antimony. As he had predicted, the temperature at the junction of the two conductors changed with the application of an electrical current. He also discovered that the temperature of the metals differed at their ends and that the current absorbed heat at one end and released it at the other. Like Seebeck, however, Peltier misinterpreted his results. With an ingeniously simple experiment—placing a drop of water at the junction of the two conductors and watching it freeze and melt depending on the direction of the current—Emil Lenz first demonstrated and correctly interpreted the Peltier effect in 1838.

From the time Lord Kelvin clarified the relationship between the Peltier and Seebeck effects in 1854 until the 1950s, research on thermoelectricity moved along at a languid pace. Bold efforts to develop practical devices based on thermoelectric principles met with little success. Known materials allowed for efficiencies of just 1 percent, far too low to justify any serious development efforts. The most important contribution to the study of thermoelectricity in the early twentieth century came from E. Altenkirch, a German scientist, who determined that progress in thermoelectricity depended on finding materials that exhibited three characteristics—(1) high electrical conductivity, (2) high voltage capacity, and (3) low thermal conductivity. Since he knew of no such materials, Altenkirch abandoned his search.

Thermoelectric researchers developed a greater understanding of the possibilities of semiconductors in the 1930s, and the positive–negative (*p–n*) junction, a crucial component of thermoelectric devices, was developed in 1942. However, not until the 1950s, after researchers at the Soviet Institute of Semiconductors declared the inevitability of a thermoelectric breakthrough and H.J. Goldsmid of General Electric's London laboratory provided a rationale for studying the heaviest semiconductor compounds such as bismuth telluride and lead telluride, was there a concerted worldwide effort to overcome the technical barriers to the development of thermoelectric devices.

At the core of technologies based on thermoelectric principles is a simple solid-state device that facilitates the exchange of thermal and electrical energy through the movement of electrons and holes (Figure 27). An electrical current passing through the device will draw heat from one side of the *p–n* junction and release it on the other side. If the current is reversed, so too is the heat transfer process. Employed in this way, as a solid-state heat pump, the thermoelectric device can be used for refrigeration, air conditioning, and heating. The

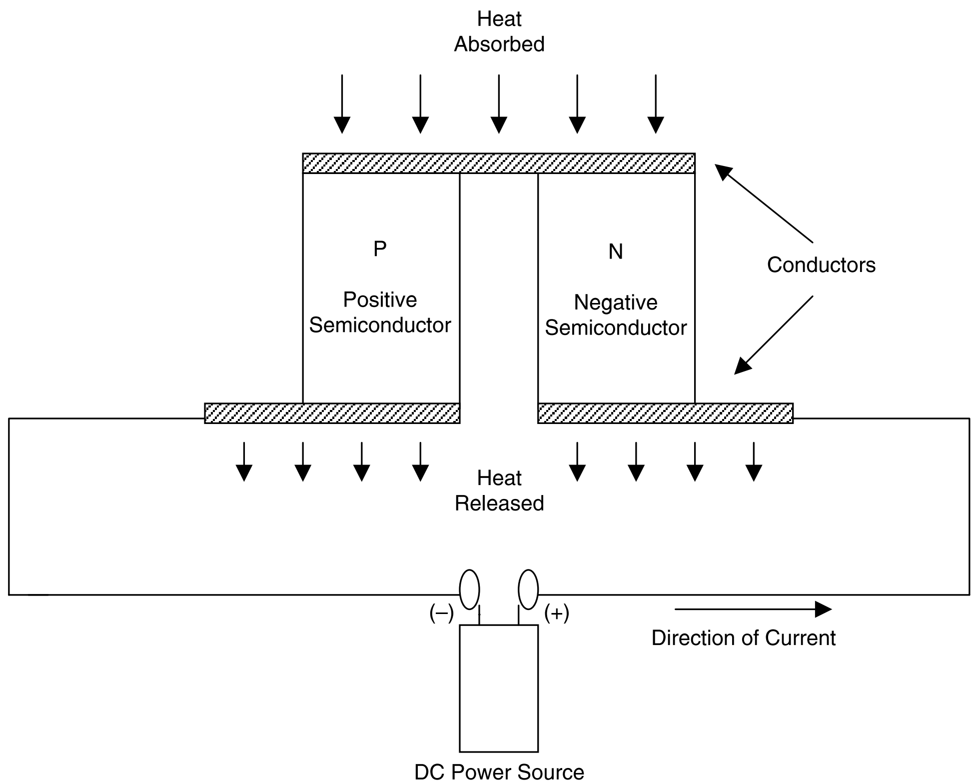


Figure 27. Schematic drawing of a simple thermoelectric device.

same device also can be used to exploit temperature differentials of opposite sides of the *p–n* junction to produce electricity. In thermoelectric generators, thermal energy is transformed into electrical energy as it passes from one side of the device to the other.

The 1950s witnessed an explosion of interest in thermoelectric research, especially for refrigeration. Optimistic about the potential of thermoelectricity, researchers in corporate laboratories convinced electrical industry executives and government research agencies to devote substantial resources to thermoelectric research and development for military and commercial applications. Researchers worldwide published twice as many papers on thermoelectricity in the four years between 1956 and 1960 as they had during the previous 130 years. In the U.S. alone, over 75 organizations, including research universities, government laboratories, private research institutes, and major corporations, maintained substantial thermoelectric research programs.

Major electrical appliance manufacturers, such as General Electric, RCA, Westinghouse, and Whirlpool, pinned their hopes on the large, growing, and highly profitable market for standardized household refrigeration units in the late 1950s. Thermoelectric refrigerators, however, could not compete in price or efficiency with existing mechanical refrigerators. Furthermore, manufacturers were unwilling to adjust their expectations and to shift their focus to smaller markets for specialty devices in which thermoelectric units might have been competitive. As the promise of big profits in the household refrigeration market evaporated by the mid-1960s, all the major appliance manufacturers and all but a handful of specialty firms abandoned thermoelectric research and development for the consumer market. Research on military applications of thermoelectricity continued nonetheless and proved productive over the long term. Thermoelectric devices have been used to power electronic equipment in spacecraft, to cool submarines, and to provide a number of other military applications in which cost and efficiency are of limited concern.

Interest in commercial applications of thermoelectricity refrigeration reemerged in the 1990s. Niche markets developed for portable coolers, and at least one automobile manufacturer embedded thermoelectric cooling units in its car seats.

Manufacturers in the U.S., Europe, and Japan began to experiment with using thermoelectric chips to cool computers and medical devices such as blood analyzers. Most of these devices used bismuth telluride alloys in simple *p–n* junctions, which were still too inefficient to employ in larger units such as household refrigerators. The technical breakthrough that researchers had been seeking for 40 years appeared to have arrived in 2002, through U.S. government-funded research begun in 1993 at the Research Triangle Institute (RTI) in North Carolina. RTI researchers used a thin film deposition process to lay down thermoelectric alloys in alternating layers. The microscopic structure, known as a superlattice, enhances electron flow while hindering heat transfer, resulting in efficiencies two-and-a-half times greater than allowed with simple *p–n* junctions. Whether devices based on superlattice structures will finally bring about the long-promised thermoelectric revolution in refrigeration and cooling remains to be seen.

See also **Refrigeration, Mechanical; Semiconductors, Compound; Semiconductors, Elemental**

GLEN ASNER

Further Reading

Finn, B.S. Thermoelectricity, in *Advances in Electronics and Electron Physics*, vol. 50. Academic Press, New York, 1980.

Goldsmid, H.J *Applications of Thermoelectricity*. Methuen & Company, London; and Wiley, New York, 1960.

Goldsmid, H.J. *Electronic Refrigeration*. Pion, London, 1986.

Joffe, A.F. The revival of thermoelectricity. *Scientific American*, 199, 5, 31–37, 1958.

Lynch, C.J. Thermoelectricity: the breakthrough that never came. *Innovation*, 28, 48–57, 1972.

Thévenot, R. *A History of Refrigeration Throughout the World*. Translated by Fidler, J.C. International Institute of Refrigeration, Paris, 1979.

Skrabek, E.A. Thermoelectric energy conversion, in *Encyclopedia of Energy, Technology, and the Environment*, vol. 4, Bisio, A. and Boots, S., Eds. Wiley, New York, 1995.

Renewable Power Generation, *see* **Biomass Power Generation; Hydroelectric Power Generation; Power Generation, Recycling; Solar Power Generation; Wind Power Generation**

Reppe Chemistry

Reppe chemistry refers to a group of high-pressure reactions used in industry to make various organic chemicals. Most of these reactions were based on acetylene, and as acetylene declined in popularity (because of the cheapness of ethylene) in the 1960s, most of the Reppe reactions have also lost their significance. However, the formation of butanediol from acetylene has survived and is now one of the few acetylene-based processes used in the chemical industry.

Walter Reppe who was working for the German firm of IG Farben at Ludwigshafen on the Rhine, discovered in 1932 that alcohols could be added to acetylene under considerable pressure to form the corresponding vinyl ether. The dangers of using acetylene were well known, and the success of this reaction was unexpected. The vinyl ethers were used as the starting point for the production of polyvinyl ethers, which were considered to be possible alternatives to polyvinyl chloride (PVC). Subsequently they turned out to have only limited applications, for instance, as a synthetic substitute for chewing gum.

Six years later, Reppe discovered a different type of reaction that preserved the triple bond of acetylene. Two molecules of formaldehyde were added to acetylene under pressure using copper acetylide, a detonator, as the catalyst. The initial product, butyne-1,4-diol was converted into butane-1,4-diol which was then used to make synthetic rubber during World War II. Since 1945, butanediol has been used to make the well-known polyurethanes and the important solvent tetrahydrofuran (THF). Reppe then attempted to make aldehydes by adding carbon monoxide to acetylene, but the reaction produced acrylic acid, a chemical that is used extensively in paint but had previously been very expensive. Learning of similar research at Ruhrchemie (the OXO process) in 1940, Reppe extended the reaction to ethylene (yielding propionic acid) and methanol (yielding acetic acid). A semi-industrial plant for the manufacture of propionic acid, used to make fine chemicals such as mold inhibitors, started up at Ludwigshafen in 1951, and a full-scale acetic acid plant followed in 1957.

In 1940, Reppe also achieved a remarkable cyclization of acetylene to the exotic compound cyclooctatetraene (COT), using a nickel cyanide catalyst. Much hope was held out for the industrial development of COT, which was not fulfilled. BASF, the postwar successor company to IG Farben at Ludwigshafen, maintained a small pilot plant for many years and provided free COT to academic researchers.

See also **Chemicals**

PETER MORRIS

Further Reading

Achilladelis, B.G. Process Innovation in the Chemical Industry. University of Sussex D.Phil. thesis, 1973, pp. 75–90.

Copenhaver, J.W. and Bigelow, M.H. *Acetylene and Carbon Monoxide Chemistry*. Reinhold, New York, 1949.

Hummel, H.G. Walter Reppe, 1892–1969. *Chemische Berichte*, 117, 1–21, 1984.

Miller, S.A. *Acetylene: Its Properties, Manufacture and Uses*, vol. 1. Ernest Benn, London, 1965.

Morris, P.J.T. The Development of Acetylene Chemistry and Synthetic Rubber by I.G. Farbenindustrie Aktiengesellschaft, 1926–1945. Oxford University D.Phil. thesis, 1982, pp. 111–117 and 123–133.

Morris, P.J.T. *Ambrose, Reppe and the Emergence of Heavy Organic Chemicals in Germany, in Determinants in the Evolution of the European Chemical Industry, 1900–1939: New Technologies, Political Frameworks, Markets and Companies*, Anthony S. Travis, et al., Eds. Kluwer, Dordrecht, 1998, pp. 89–122.

Morris, P.J.T. An industrial pioneer: Walter Reppe. *Chemi. Br*., 29, 38–40, 1993.

Morris, P.J.T. The technology–science interaction: Walter Reppe and cyclooctatetraene chemistry. *Br. J. Hist. Sci*., 25, 145–167, 1992.

von Nagel, A. *Äthylen, Acetylen*. Firmenarchiv der BASF, Ludwigshafen, 1971.

Reppe, W. *Neue Entwicklungen auf dem Gebiete der Chemie des Acetylen und Kohlenoxyds*. Springer, Berlin, 1949.

Reppe, W. *Chemie und Technik der Acetylen-Druck-Reaktionen*. Verlag Chemie, Weinheim, 1951; expanded edition, 1952.

Stokes, R.G. *Opting for Oil: The Political Economy of Technological Change in the West German Chemical Industry, 1945–1961*. Cambridge University Press, Cambridge, 1994, pp. 35–39.

Research and Development in the Twentieth Century

By the end of the nineteenth century, the emergence of large multinational industrial companies had begun. Along with this growth, companies established different departments for each specific aspect of the industrial process. This is what Alfred Chandler called the "visible hand" in industrial management in that period, and he showed that this phenomenon took place in countries such as the U.S., Britain, and Germany. One of the tasks that became the main focus for a separate department was the research and development (R&D) function. In various countries, large companies set up laboratories to conduct R&D that would create new products or improve existing ones. The first sector in which such laboratories emerged was the chemical industry, soon followed by the electrotechnical sector. In the U.S., for example, General Electric (GE) started a central laboratory in 1900, DuPont in 1902, and Eastman Kodak in 1910. In Germany, the Siemens company set up its main laboratory in 1905; in the U.K., the General Electric Company (GEC) started a central laboratory in 1919; and in The Netherlands the Philips Company initiated its *Natuurkundig Laboratorium*, or "physics laboratory," in 1914.

One of the driving factors for R&D was the redefinition of patent laws. In the U.S. this was the 1890 Sherman Antitrust Law. In order to assure that industries were free to act and to maintain a sound position in the market, companies had to develop their own knowledge bases. This was one of the tasks of the new laboratories.

Research and development in the twentieth century will be discussed mainly for the chemical and electrotechnical industries because they are knowledge-intensive and have therefore traditionally received most attention in terms of their R&D. Among other areas that have also become knowledge intensive is the pharmaceutical industry. Considerable developments have taken place through R&D; for example, vitamins in the 1930s, steroids, antibiotics, and cardiovascular products in the 1940s and 1950s. The area of materials can also be mentioned as one in which much R&D was undertaken. However, much less has been written about the history of R&D in these sectors than for the chemical and electrotechnical sectors.

In many companies with a new central laboratory, the first effect of the lab activities was the diversification of the company's product portfolio. This can be seen clearly with the examples of GE and Philips. Both companies were started at the end of the nineteenth century to produce light bulbs only. However, as their laboratories started undertaking research to study phenomena related to the functioning and production of light bulbs, they soon realized that the same phenomena also accounted for the functioning of other devices, such as x-ray and radio vacuum tubes. As a result, increased insight into the light bulb phenomena also contributed to the development of other products. Usually there was close contact with the company's management so that the decision to diversify the product portfolio was often a matter of cooperation between the company managers and the lab managers. When both GE and Philips had moved into the market of radio vacuum tubes, they took the next step in the diversification process by extending R&D activities to other elements of the radio: electric circuits, loudspeakers, and later sound recording. In the beginning all these components were treated independently. This can be called "device research." Later, the

integration of these parts into a whole (e.g., a complete radio set, or a complete radio broadcast and receiving system) became a separate concern for research activities. This was the beginning of what later would be called "systems research" as a third type of research alongside materials and devices research. Most of these developments took place in the first half of the twentieth century.

Although World War II had a large impact on many aspects of social life, it did not seem to have much influence on the research and development activities of most large companies. Even in The Netherlands, which was occupied by the Germans from 1940 to 1945, the research program of the Philips Company to a large extent was continued. Some activities were terminated because the outcomes might be useful for military use by the Germans, and others were continued secretly. In countries that the Germans did not occupy, such as the U.K. and the U.S., World War II stimulated research for military purposes. The main impact of the war, however, was the successful development of the atomic bomb, which created an awareness that research into fundamental phenomena could result in dramatic breakthroughs. This caused the American president's scientific advisor, Vannevar Bush, to produce a report titled *Science, the Endless Frontier*. The main message of this report was that basic research was to be stimulated by the government and be taken up in industrial companies, because in the end it would always lead to industrial applications, often of a breakthrough character. This claim was soon supported by another important example, namely the invention of the transistor at the Bell Labs in 1947. This invention took place in a research group that fitted well with the ideas that Vannevar Bush had formulated. Earlier efforts to make a transistor by creating a solid-state analog of the triode had failed, and it was not until solid-state physics was applied to the problem that the transistor was invented. Not long afterward, the laser was invented; this too was a result of the application of solid-state physics, thus confirming the Vannevar Bush philosophy of basic research resulting in industrial application.

The ideal of basic research led to a new role for the research laboratories of the industrial companies. A substantial part of R&D was now dedicated to research into phenomena without any concrete relationship to possible applications. Sometimes, as in the case of the Philips Company, this change was accompanied by establishing development laboratories in the product divisions. Thus basic research was conducted in the central laboratory, while the division labs undertook applied research. The basic research approach did yield some great successes; two discoveries at Philips illustrate this point. First, the Plumbicon, a television pickup tube, was invented in 1958, and it would replace earlier tube types in all professional television cameras. Second, Philips researchers invented local oxidation of silicon (LOCOS), a technique for producing very compact integrated-circuit (IC) structures in silicon. Both were the outcome of research that was oriented to fundamental phenomena, and the patents on these inventions created a large income for the company as competing companies realized that they were almost forced to use these inventions to maintain their position on the market. The field of ICs is related to another trend, the internationalization of R&D activities from 1950 to 1970. The Philips Company, for example, established research laboratories in the U.K., France, Germany, and the U.S. in the first decade after World War II. There were regular contacts between the research laboratories of the various companies. This sometimes led to important exchanges of knowledge; for example, the Bell Labs transistor knowledge was made available to Philips based on the fact that Philips previously had made available to Bell certain knowledge of ferrites (nonmetallic magnetic materials). In the last decades of the twentieth century, globalization of R&D increased rapidly. For example, R&D expenditures by foreign-owned companies in the U.S. more than doubled, from \$6.5 billion in 1987 to \$14.6 billion in 1993. By 1999 this amount had further increased to \$17 billion.

To be successful in basic research, the research laboratories claimed they needed a large degree of independence and freedom. In a number of companies the research laboratories received their budgets directly from the company's management, which made them financially independent of the product divisions. The result was that product divisions had almost no say in the research program. At the same time, the product divisions felt no obligation to adopt research output for which they had not paid. This mutual lack of commitment in companies with separate research facilities often frustrated cooperation between research and development in this period. A striking example of this was the research on the hot-air engine that for many years took place in the Philips *Natuurkundig Laboratorium*. The origin for this research had been the need for a small, silent energy source for radio transmitters and receivers in developing countries. This need, however, had

fallen away after the transistor had replaced vacuum tubes in radios, and a simple battery would do for the energy supply. However, the Philips researchers decided to continue the hot-air engine research with car engines as a new application. There was not a single product division in the company for which that would have been a useful development. The financial independence of the research lab allowed it to continue the work on its own. It never yielded any profit, and it was not terminated until an entirely new situation for the research lab was established in the late 1970s.

In some companies, DuPont for example, the labs in the various company divisions also undertook a certain amount of long-term basic research. In this company, such research in earlier years had led to the development of nylon. The hope was that basic research in the division labs would result in finding similar new products. When this expectation was not fulfilled, the company management began to suspect the validity of basic research promises. The economic decline of the late 1960s forced companies to critically review the basic research of central laboratories. In this review it became clear that the number of major breakthroughs due to basic research in the 1950s and 1960s had been fairly small compared with the effort that had been spent on this type of research. This result was found not only at DuPont but also in electronics companies such as Philips, GE, and RCA.

In addition to economic factors, increasing social criticism of technological developments caused companies to reconsider their research and development activities. The first report from the Club of Rome in 1972 had placed concerns about environmental damage by technological developments on the political agenda. Together with increased doubts about the value of basic research in industrial companies, these concerns would lead to a new position for R&D within companies. Society, not science, was to be the new "Endless Frontier" (a term from the 1988 European Commission report, titled *Society, the Endless Frontier*).

The transition to the new situation can be suitably illustrated by the work that was undertaken on optical recording in various companies. At both Philips and RCA, the research labs led this technological development. The concrete outcome of this effort, the videodisc, was a commercial failure for all companies that were involved in the field. The same knowledge was later used at Philips and Sony to develop the compact disk (CD). In that development, however, the product divisions had the leading role and the research labs were called in "only" for delivering specific knowledge on optical recording. This would establish the new relationship between research and development in such companies: the division in which the technological development work took place would decide on the desirability and feasibility of new products, and the research laboratories were commissioned to deliver the specific knowledge to be used by the divisions to develop the products. The idea of basic research leading to development work on new products was abandoned. This new approach also led to an almost complete "divisionalization" of research programs; only a very small amount of basic research was kept.

Market requirements became increasingly important for both research and development activities. In the 1950s and 1960s, economic growth enabled companies to market all sorts of new products, and rarely was the introduction of a new product a commercial failure. However, the economic decline in the early 1970s made customers aware that they could only spend their money once, and they became more critical of new products. Not only were they more hesitant to buy every new gadget that appeared, they were also more critical in comparing products offered by competing companies. Taking into account the customers' desires therefore became a crucial issue for product development. This led to a new way of considering the concept of quality. In the past, quality had mainly been a concern for production departments, and it generally meant that the output of production lines was checked for obvious defects. With increased attention to customers' wishes, however, came the awareness that the customer had to be taken into account not only in the production phase but also in the development phase, and perhaps even in the research phase. Quality was to be "built in" to the design of the product. For that purpose, new methods for supporting product development were created. The term total quality management (TQM) was often used to describe the overall effort of assuring quality throughout the development and production activities. It became evident in time that the application of such methods presented other problems. In general, the concept of quality management had previously been used in the context of production activities that had a repetitive and predictable nature. Often statistics played a role in such methods. R&D activities were much less repetitive and predictable, and this created a tension between quality methods and the nature of such activities. Yet quality management clearly found its way into the development work of

industrial companies, and even in research laboratories, quality managers were appointed to stimulate and monitor quality-oriented efforts. For researchers, particularly those who had been used to a large degree of freedom for their work, this was a significant change in culture.

Quality methods helped product developers take into account the whole life-cycle in the development of new products and try to create designs such that all phases would go smoothly. This resulted in a collection of methods called ''design for *x*,'' where *x* could be any phase in the life-cycle of the product: ''design for manufacturing,'' ''design for assembly,'' ''design for logistics,'' ''design for disassembly,'' ''design for recycling,'' and so on. A method was developed in Japan that specifically focused on the customer's requirements and was called quality function deployment (QFD). In general, Japanese companies were the first to take the customer's requirements into account, while Western companies still focused on technical perfection. One example of the success of the customer-oriented approach in the marketplace is seen in the competition between the digital compact cassette (DCC) that Philips released in the late 1980s and the minidisc that Sony brought out almost simultaneously. Philips had directed its efforts toward the technical operation of the product, while Sony had carefully looked at possibilities for attracting the customer's attention by adding certain functions. For example, the minidisc was almost immediately produced in a portable version, and unlike the DCC, the user could program the title of a music track into the disc so that it would be displayed while playing. Striving for technical perfection had also caused Philips to abandon the plan to release the DCC before Sony could bring out its minidisc. The result was that both came on the market almost simultaneously, allowing customers to compare them and recognize the advantages of the minidisc. Such stories made it clear that three issues were crucial for success in R&D in the 1980s and 1990s: cost, time-to-market, and quality (in the sense of customer satisfaction).

A second trend that related to the need to consider the product's life-cycle in industrial R&D was the concern for environmental effects of technological developments. This concern became a political issue in the early 1970s, and by the 1980s and 1990s it was ''operationalized'' into guidelines for industrial activities, not only in production but also in R&D. Here, too, the idea was that environmental concerns should be built into the design for new products. Rather than rejecting polyvinyl chloride (PVC) because it caused polluting emissions in the production phase, researchers would reconsider its merits and take into account the fact that it could be recycled. To make a balanced, overall evaluation of the environmental impact of a product, the ''life-cycle analysis'' (LCA) was developed as a new method in product development. In an LCA, the sum of a number of environmental effects for all phases in the life-cycle is presented. Such effects may include the contribution to ozone layer depletion, contribution to the greenhouse effect, and contribution to solid waste. LCAs allow designers to compare different materials and processes with respect to their impact on the environment. Even though there are several methodological problems, as in the field of quality methods, LCAs have found their way into industrial practice.

The changes in approach to R&D in the twentieth century have sometimes been described in terms of generations. All R&D before the 1970s is called first generation R&D, and technology is seen as the main asset in this generation. Second generation R&D is related to the rest of the business functions, as before World War II, and attention is shifted toward individual projects as the main asset. By the 1980s R&D sought to invade the whole enterprise more systematically. Fourth generation R&D means that learning from the customer becomes crucial in the R&D function. The newest trend at the end of the twentieth century was called fifth generation R&D, where knowledge in general is regarded as the main asset. Of course, different companies moved through these generations in different ways, but they do give an overall indication of the dynamics of R&D in the late twentieth century.

MARC J. DE VRIES

Further Reading

Bralla, J.G. *Design for Excellence*. McGraw-Hill, New York, 1996.

Chandler, A.D., Jr. *Scale and Scope. The Dynamics of Industrial Capitalism*. Belknap Press of Harvard University Press, Cambridge, MA, 1990.

Clayton, R. and Algar, J. *The GEC Research Labs, 1919–1984*. Peter Peregrinus, London, 1989.

Graham, B.W. *The Business of Research. RCA and the Video Disc*. Cambridge University Press, Cambridge, 1986.

Hounshell, D.A. and Smith, J.K., Jr. *Science and Corporate Strategy: DuPont, 1902–1980*. Cambridge University Press, Cambridge, 1988.

Reich, L.S. *The Making of American Industrial Research. Science and Business at GE and Bell, 1876–1926*. Cambridge University Press, Cambridge, 1985.

Rosenbloom, R.S. and Spencer, W.J., Eds. *Engines of Innovation: US Industrial Research at the End of an Era*. Harvard Business School Press, Boston, 1996.

Stokes, D.E. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press, Washington D.C., 1997.

De Vries, M.J. *80 Years of Research at the Philips Natuurkundig Laboratorium (1914–1994). The Role of the Nat. Lab. at Philips*. Stichting Historie der Techniek, Eindhoven, 2001.

Rocket Planes

Once the Wright brothers had proved that controlled and sustained powered flight was possible, two related avenues of development were immediately apparent: increasing the distance and the speed of flight. From that first flight in 1903 until the late 1930s, aircraft control and propulsion technology was developed to improve all aspects of flight, culminating in the production of the subsonic turbojet. Although further development of the jet engine would revolutionize both military and commercial aviation and boost speeds into the supersonic region, what became known as the "sound barrier" (Mach 1) was first broken by a rocket-powered aircraft, the XS-1.

Subsonic World War II aircraft had approached Mach 1 in steep dives, but the results were often disastrous since the materials and control systems of the time were unable to cope with the aerodynamic forces encountered. Part of the solution was to ensure that an aircraft passed through the transonic zone as quickly as possible, which meant providing thrust levels higher than the jet engines of the time could deliver. The obvious solution was the rocket engine.

A number of experiments in rocket-assisted flight were conducted in Germany from 1928, when experimenter Friedrich Stamer flew 1.2 kilometers in a small glider propelled by an elastic rope and two small rockets. In 1929, a Junkers 33 seaplane made a rocket-assisted take-off at Dessau, Germany, and by the late 1930s the Heinkel 176, the first rocket-powered aircraft designed for sustained flight, had been tested successfully. This led to the development of the Messerschmitt Me-163 research plane, which first flew in 1940 and set a new world speed record of 917 kilometers per hour (km/h) in 1941. This attracted the attention of the military authorities, who proceeded to develop the Me-163 as a fighter, renaming it Komet. More than 360 were built and some were used effectively in World War II, but by then rocket research had switched to the development of ballistic missiles, the most notable result of which was the V-2 "vengeance" weapon, which became the technical basis for both Soviet and American rocket programs after the war.

Meanwhile in the U.S., the need for similar research and development was being debated. In 1943, the National Advisory Committee for Aeronautics (NACA, the forerunner of the National Aeronautics and Space Administration, NASA) conceived a rocket-propelled research aircraft designed to exceed Mach 1 in horizontal flight. Thus on October 14, 1947, pilot Charles E. (Chuck) Yeager exceeded the speed of sound in the XS-1 (Experimental Supersonic 1). Between them, the XS-1 and two later planes designated X-1, made 156 flights, setting a speed record of 1540 km/h (about Mach 1.45) and an altitude record of 21,916 meters. The aircraft, which were dropped from the modified bomb-bay of a B-29 bomber for their test-flights, were designed and built by Bell Aircraft Corporation and were sometimes known as the Bell X-1. They were powered by a Reaction Motors rocket engine, burning ethyl alcohol and liquid oxygen, which produced about 2700 kilograms of thrust.

The aircraft themselves were designed to have unprecedented structural strength to survive the aerodynamic forces of the transonic region (up to 18 g (the acceleration due to gravity) and –10 g), and carefully shaped to reduce drag. By reducing the cross-sectional area of the fuselage toward the rear, the air compressed by the plane's motion and forced back along the fuselage in a shock wave would have a chance to expand slightly, thus easing the aircraft's progress through the sound barrier. The resulting improvement in aerodynamic efficiency meant that engines generating much lower thrust levels could achieve supersonic speeds (Figure 28).

Later X-planes were designed to go higher and faster than the X-1 and to test new materials that would survive the frictional heating effects of ever-higher velocities. This culminated in the development of the X-15, a rocket-propelled research aircraft designed and built by North American Aviation to investigate the effects of high speed at high altitude on future space planes (Figure 29). The main engine burned the propellants, anhydrous ammonia and liquid oxygen, while the altitude control thrusters used for steering consumed hydrogen peroxide. NASA's fleet of three X-15s was dropped from beneath the wing of a converted B-52 bomber in a series of 199 flights (the first and last powered flights occurred in September 1959 and October 1968, respectively). Designed to reach speeds of up to Mach 6 in

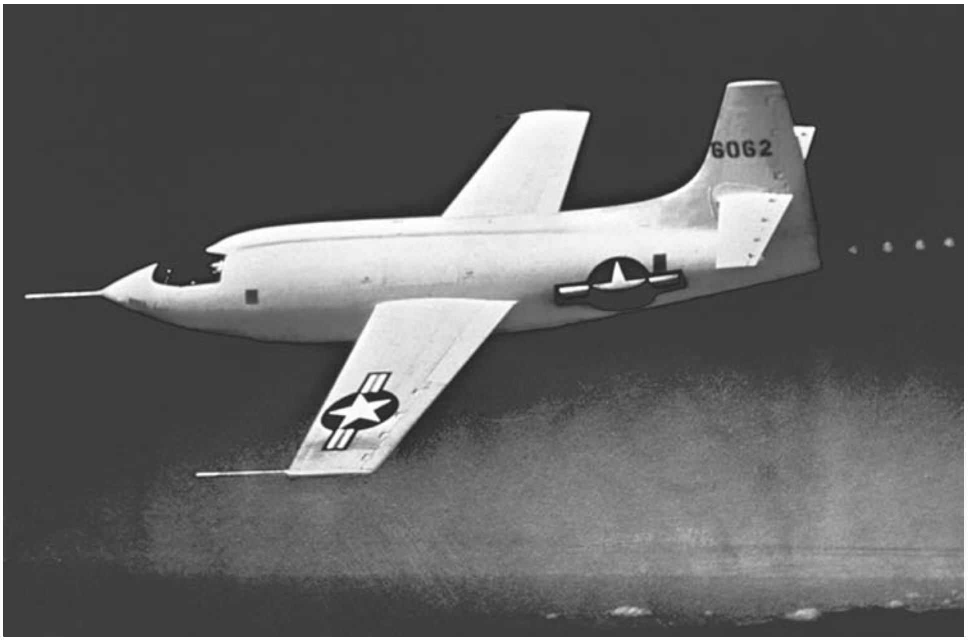


Figure 28. The Bell Aircraft Corporation X-1-1 in flight, showing a shock wave pattern in its exhaust plume. This aircraft was nicknamed "Glamorous Glennis" by Chuck Yeager in honor of his wife, and is now on permanent display in the Smithsonian Institution's National Air and Space Museum in Washington, D.C.
[*Photo courtesy of NASA*.]

Figure 29. An X-15 rocket plane after landing. As ground support personnel begin post-flight activities, the NASA B-52 launch aircraft and two F-104 chase aircraft provide an aerial salute.
(*Photo courtesy of NASA*.)

horizontal flight, the X-15 set a record of 7274 km/h (about Mach 6.7) and, separately, an altitude record of 107,960 meters. One of its more famous pilots was Neil A. Armstrong, the first man to walk on the moon.

Despite their success as research and technology demonstrators, rocket planes turned out to be a technological cul-de-sac. Interest in rocket-powered aircraft waned as the performance of jet engines improved to the point where, less than ten years after Yeager's historic flight, jet-powered fighters capable of Mach 2 were production-line items. Initial plans to fly the X-15 into orbit were cancelled in favor of vertically launched rockets, which were to become the preferred delivery system for NASA's Mercury program and those that followed. Nevertheless, X-15 research did engender an extensive program of winged lifting-body research, which led eventually to the Space Shuttle. Moreover, it is widely believed that expendable rockets will one day be superseded by reusable space-planes.

See also **Aircraft Design; Rocket Propulsion, Liquid Propellant; Space Launch Vehicles**

MARK WILLIAMSON

Further Reading

Caidin, M. *Wings into Space*. Holt, Rinehart & Winston, New York, 1964.

Godwin, R. *X-15: The NASA Mission Reports*. Apogee Books, Ontario, 2000.

Johnston, A.M. and Barton, C. *Tex Johnston: Jet-Age Test Pilot*. Smithsonian Institution Press, Washington D.C., 2000.

Rotundo, L. *Into the Unknown: The X-1 Story*. Smithsonian Institution Press, Washington D.C., 1994.

Yeager, C. and Janos, L. *Yeager: An Autobiography*. Century, London, 1985.

Rocket Propulsion, Liquid Propellant

The world's first rocket, probably originating in China in the eleventh century AD, was solid-propelled with ordinary gunpowder as the propellant. Eight hundred years later, from the late nineteenth to the early twentieth centuries, Konstantin Tsiolkovsky of Russia, Robert H. Goddard of the U.S., and Hermann Oberth of Transylvania independently conceived the idea of the liquid-propellant rocket. These pioneers made the important theoretical discovery that the rocket operates according to Isaac Newton's third law of motion: "For every action there is an opposite and equal reaction." This means that reaction propulsion operates both on the ground and in the vacuum of space. This was in total contradiction to the widely held but erroneous belief that the rocket works because the gases push against the air. The pioneers went further and realized that liquid propellants were potentially a far more efficient way to achieve space flight for several reasons.

Liquid propellants consist of a fuel (the substance burned) and the oxidizer (the liquid containing oxygen atoms needed for combustion of the fuel). The liquids are kept in separate tanks and pumped into a combustion chamber at regulated pressures by pumps, in a process of controlled combustion. The two ingredients burn in the chamber and expel their gases out of the rear opening of the rocket, driving the rocket forward by reaction propulsion. Using valves, liquid propellants can be stopped and started at will. An important extra advantage with liquids is that they are more powerful than solids. All three of the pioneers found that the combination of liquid oxygen (oxidizer) and liquid hydrogen (fuel) was the ideal propellant, although these liquefied gases are extremely cold (below 0°C) and were then very difficult to manufacture. For this reason, Goddard, the only one of the pioneers to undertake extensive experiments, chose to use readily obtainable but less powerful liquid oxygen ("LOX") and gasoline. From 1921 until his death in 1945, Goddard engaged in liquid-propellant rocketry experiments.

On March 16, 1926, Goddard launched the world's first liquid-fuel rocket at Auburn, Massachusetts. It went up to 12.5 meters in 2.5 seconds, but this effort was not publicized. Earlier, Tsiolkovsky first published his theories of the liquid-propellant (LOX/hydrogen) space rocket in 1903 in a Russian popular science magazine. Although he continued to publish his theories, his work was almost unknown in the West. In 1923, Oberth's seminal book, *Die Rakete zu den Planetenräumen [The Rocket into Planetary Space]* appeared. It was enormously influential and virtually created a space flight and rocket fad. Oberth described several models of LOX/alcohol rockets in detail, including pumps and methods of cooling, and the possibility of manned space missions.

As a result, several amateur rocket groups were formed around the world. The most prominent was the *Verein für Raumschiffahrt*, or VfR (Society for Space Travel). They began crude experiments with liquid-propellant rockets from 1930. The German Army began its own secret rocket program and in 1932 hired Wernher von Braun, a young and gifted VfR member to be the technical director. By 1940,

the German Army had developed the A-4 (Aggregate 4) rocket, later known during World War II as the V-2. Using LOX and alcohol as propellants pumped into an hourglass-shaped combustion chamber with 18 separate propellant injectors on its dome, the V-2 produced 27,000 kg of thrust with a range of 320 kilometers. Cooling was by a combination of regenerative cooling, in which the fuel flowed around a cooling jacket surrounding the chamber before it was injected from the top, and "film cooling," in which the fuel was sprayed onto the inside of the nozzle wall, resulting in a thin film of cool fuel. The V-2 was the world's first large-scale liquid-fuel rocket. However, despite a popular misconception, V-2 development was more influenced by Oberth than Goddard. Under self-imposed secrecy, Goddard worked in the New Mexico desert and achieved remarkable results, notably gyroscopically stabilized, pump-driven rockets and thrusts of almost 450 kilograms. The highest altitude in his flight tests was 2740 meters on March 26, 1937. Goddard was unaware that the Germans had attained three-dimensional gyroscopic control and thrusts of 1485 kilograms pounds with their A-3 test rocket, which was not meant for altitude flights. The Germans also developed other missiles.

In 1941 Reaction Motors, Inc. (RMI) was formed by four American Rocket Society (ARS) members and was based on the success of James H. Wyld's regeneratively cooled rocket motor design. Similar to the V-2 design, the fuel in the Wyld motor circulated around a cooling jacket before injection. RMI adapted this technique to JATOs (jet-assisted-take-off) rockets for heavily loaded seaplanes and to small, experimental missiles. Following the war, both the U.S. and the Soviet Union captured German V-2 rockets, acquired V-2 technicians, and began developing their own long-range missiles. In 1946 the North American Aviation Company initiated development of its 800-kilometer-range Navaho. Navaho was eventually upgraded to an intercontinental missile, and it established the most important milestones in the history of U.S. rocket technology.

At first, the Navaho was designed as a modified V-2 with wings. When the Air Force requested a change to a 1600-mile-range weapon, the Navaho was designed as an air-breathing ramjet-powered cruise missile with large liquid-propellant boosters. To learn how to build and handle large rocket engines, North American made three copies of the V-2 engines. By 1950 they had created an entirely new and more compact cylindrical, single flat-plate injector engine, the XLR-43-NA-1. Still using LOX/alcohol, this engine produced 34,000 kilograms of thrust. This was the first large-scale liquid fuel engine developed by the U.S.

Also in 1950, the U.S. Army, working with the emigrant German scientist Wernher von Braun, began development of its 800-kilometer-range Redstone missile, essentially a super-V2. In the following year, when Navaho was further upgraded and required more powerful boosters, the Army adopted the XLR-43 for the Redstone. The Redstone subsequently used a modified XLR-43 and became the U.S.'s first operational surface-to-surface ballistic missile. Later, with a more powerful hydrazine-based propellant and producing 37,700 pounds of thrust plus upper stages, the modified Redstone (also known as the Jupiter C) launched the U.S.'s first artificial satellite Explorer 1 on January 31, 1958. The Redstone also launched America's first astronaut, Allan B. Shephard, into space on May 5, 1961, on a suborbital flight.

Meanwhile, the North American Aviation Company produced the more powerful XLR-43-NA-3 for the Navaho on November 19, 1952. It was the first U.S. liquid-fuel rocket engine to reach a thrust of more than 100,000 pounds (45,360 kilograms), with a maximum thrust of 120,000 pounds (54,430 kilograms). Among the engine's major technological breakthroughs was the "spaghetti" or tubular form of combustion chamber, in which the cooling tubes formed the walls of the chamber itself. This reduced the weight of the chamber by 50 percent. The spaghetti configuration became adopted throughout the U.S. rocket industry for large-scale engines such as the Space Shuttle main engine (SSME). Edward A. Neu, Jr., of RMI conceived the idea about 1947, but other companies also developed the concept.

In 1952, the U.S. Air Force Rocket Engine Advancement Program (REAP) saw another important development applied to the Navaho and later engines, the use of LOX/hydrocarbon propellants like JP-4, or kerosene. Hydrocarbon fuels replaced V-2 era alcohol and increased the specific impulse, a unit of measurement for the efficiency of rocket engines. North American used the regeneratively cooled, tubular chamber with flat-head injector and LOX/JP-4 as a building-type engine. In the upgrade of the Navaho, two of the engines were coupled together to produce a 240,000-pound (108,862-kilogram) thrust engine, tested at full thrust in 1954.

The rapid success of North American's large-scale engine development led to contracts to develop similar engines for the Atlas intercontinental ballistic missile (ICBM) and the 2400-kilo-

meter Jupiter and Thor intermediate range ballistic missiles (IRBMs). In the Jupiter program, another breakthrough was made. G.V.R. Rao, a mathematician at Rocketdyne (a division of American Aviation, later Rockwell), found that a bell-shaped nozzle theoretically provided optimum thrust over standard conical nozzles. Rocketdyne successfully tested this design in January 1956 with a Jupiter engine. The bell-shaped chamber became another industry-wide feature in large-scale U.S. rocket engines.

Navaho first underwent flight testing in November 1956, but several of the vehicles failed. On July 11, 1957, the program was cancelled, mainly due to its enormous cost of almost a billion dollars. These medium-range ballistic missiles nonetheless left an enormous technological legacy in rocket propulsion. From its building-block engine derived directly from Navaho, beginning in 1959 Rocketdyne developed the F-1 engine, which powered the first stage of the three-stage Saturn-V launch vehicle used in ten Apollo missions. This single-chamber liquid-fuel engine with 680,000-kilogram thrust was the largest and most powerful single liquid-fuel engine ever utilized in rockets. The Space Shuttle main engine was part of the same lineage. (The use of LOX/hydrogen in rocket engines was pioneered by the Pratt & Whitney Company, which developed the Centaur high-energy upper-stage engine. Rocketdyne adapted this technology to their H-1 and J-2 engines for the Saturn-1B and Saturn-V, respectively.)

In other countries, notably the former USSR, the development of the liquid-fuel rocket evolved along quite different paths. Following World War II, the Soviets also acquired captured V-2 hardware and German technicians. However, the Soviets, principally under the technical leadership of Sergei P.Korolev, used V-2 variant designs in rocket engines and vehicles for a far longer period than the Americans. The Soviets did not make the same breakthroughs in lightweight engines but used such techniques as the "cluster of clusters" approach; that is, their large rockets utilized many engines firing together to achieve large thrusts. These rockets, such as the Vostok series, were thus extremely large and inefficient for their size, but they did function. A Vostok vehicle enabled the Soviets to place the first artificial satellite, Sputnik-1, into orbit on October 4, 1957, thereby opening the space age.

Liquid-propellant rocket engines have been used for several applications including sounding, or upper atmospheric research, rockets, rocket-sled propulsion, rocket research aircraft, and small attitude control and course correction thrusters for spacecraft. The latter are the simplest types of motors, often using single or monopropellants, pressure-fed systems rather than pumps, and high heat-resistant walls for cooling.

See also **Missiles, Long Range Ballistic; Missiles, Surface-to-Air and Anti-Ballistic Missiles; Space Launch Vehicles**

FRANK H. WINTER

Further Reading

Baker, D. *The Rocket*. Crown Publishers, New York, 1978.
Bilstein, R.E. *Stages to Saturn*. National Aeronautics and Space Administration, Washington D.C., 1980 and 1996 editions.
Dornberger, W. *V-2*. Viking Press, New York, 1954.
Goddard, E.C. and Pendray, G.E. Eds. *The Papers of Robert H. Goddard*. Macmillan, New York, 1970.
Smith, M. *Space Shuttle*. Haynes, Sparkford, 1985.
Uhl, M. *Stalins V-2*. Bernard & Graefe-Verlag, Bonn, 2001.
Winter, F.H. *Rockets into Space*. Harvard University Press, Harvard, MA, 1990.

Rocket Propulsion, Solid Propellant

Solid propellants typically consist of a mixture of fuel and oxidizer, whereas liquid-propellant components are generally stored in separate tanks and combined in a combustion chamber. Although solid and liquid rocket propellants are usually employed separately, they are used together in the hybrid rocket, the most common combination being solid fuel and liquid oxidizer. The propellants used in the cold-gas thrusters employed to position some spacecraft in orbit are unusual in that they are generally stored as a gas.

The rocket in its simplest form is believed to have been invented by the Chinese in the eleventh century AD using a solid propellant similar to what was later called gunpowder. The Englishman William Congreve initiated modern rocketry research and development in the early nineteenth century when he built the first gunpowder rocket weapon systems to challenge Napoleon's invasion plans for England. Solid rockets continued to be developed throughout the nineteenth century and were used during World War I, but their power and efficiency (now measured by the quantity "specific impulse") were limited by the available propellants. Thus it was that Russian rocket pioneer Konstantin Tsiolkovsky suggested the use of liquid propellants in 1903, and American researcher Robert Goddard began their development in the early 1920s. Goddard had conducted

tests with solid fuels to refine his technique, and in 1915 he found that a tapered nozzle increased the expulsion efficiency of the exhaust gases by some 64 percent. He subsequently used the nozzle in all his rocket work, and it was adopted by others. Since the 1920s, the majority of rocket systems required to carry significant payloads, especially into space, have been based on liquid propellants. Despite this, solids have their uses.

By the end of the twentieth century, solid-propellant rockets were available in all shapes and sizes and were often chosen over liquid systems because they are simpler and can be stored long-term without maintenance. They are widely used, for example, by the military services to power missiles, torpedoes, and aircraft ejection seats. On a larger scale, their inherent storability has made them ideal for intercontinental ballistic missiles (ICBMs), which may be required at very short notice. Solid-propellant rockets are also used in the space industry as launch vehicle stages, strap-on boosters and stage separation devices, and sometimes as "upper stages," which boost spacecraft to their final orbit or toward the planets.

There are two main types of solid propellant: homogeneous or double-base propellants, which are limited in power (e.g., nitrocellulose plasticized with nitroglycerin plus stabilizing products); and the heterogeneous or composite type. Composite propellants consist of a mixture of fuel and oxidizer, the latter providing the oxygen for combustion. An oxidizer in common use is crystalline ammonium perchlorate (NH_4ClO_4 or AP). It is mixed with an organic fuel, such as polyurethane or polybutadiene, which also binds the two components together. The inclusion of a plastic binder produces a rubbery material, making the propellant relatively easy to handle. Performance is enhanced by adding finely ground (10 micrometers) metal particles of aluminum, for example, which increase the heat of the reaction due to the formation of metal oxides. As an example, the propellant used in the Space Shuttle solid rocket boosters is comprised of 14 percent polybutadiene acrylic acid acrylonitrile (binder and fuel), 16 percent aluminum powder (fuel), 69.93 percent ammonium perchlorate (oxidizer), and 0.07 percent iron oxide powder (catalyst).

A propulsive device that uses solid propellants to move a vehicle is known as a rocket motor or solid rocket motor (SRM). Although the terms motor and engine are interchangeable in colloquial English, it is customary to call propulsive devices using a solid propellant motors, and those using a liquid propellant engines.

A typical SRM comprises only a few major components: a motor case that contains the propellant grain (or charge), a surrounding insulating blanket or propellant liner, an ignition system, and an exhaust nozzle (Figure 30). The motor case is typically made from a carbon composite or Kevlar composite by a process known as filament winding, which produces a strong yet lightweight structure. The propellant liner acts as both a thermal insulation blanket and a flame inhibitor and supports the propellant grain during manufacture, as it is poured into the open-ended motor case.

A typical ignition system operates as follows: a power supply sends a current pulse to a pyrotechnic cartridge, which ignites a small sample of the propellant in a steel or glass-fiber housing. This produces a controlled amount of hot gas that ignites the main motor, much like a detonator in an explosive device. The typical exhaust nozzle, or

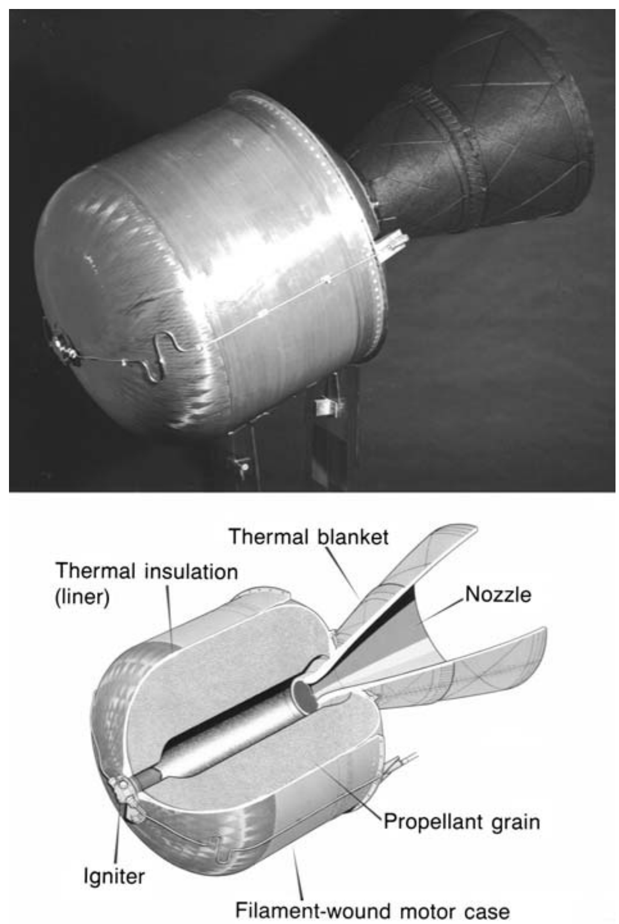


Figure 30. A solid rocket motor used as a satellite apogee kick motor; the diagram shows the major components. [*Reproduced with permission from Snecma.*]

expansion nozzle, consists of two sections: one convergent, the other divergent. The narrowest part of the nozzle, called the throat, is designed to maintain the required pressure within the motor case and regulate the outflow of combustion gases; it is also the region of transition from subsonic to supersonic flow. The exhaust gases expand and accelerate rapidly as they leave the motor, and it is the exit cone that controls the expansion of the exhaust plume.

In operation, the SRM has no need of the complex piping, pumps, and pressurization systems of the liquid rocket engine. The SRM must, however, be designed to give precisely the required amount of thrust and total impulse since it cannot be throttled, easily stopped, or restarted. Where two SRMs are used together, as on the solid rocket boosters of the Space Shuttle, they must also be precisely matched to provide the same thrust profile and to ensure that they cease firing together.

Since the thrust available from a block of propellant is proportional to the area of the combustion surface, the only way to control the magnitude of the thrust is to offer varying surface areas for combustion throughout the burn. The active area can be increased by making a cylindrical hole through the center of the block so that the hole enlarges radially and the thrust increases as the propellant burns. Constant thrust, which is more often required, can be provided by a star-shaped bore (Figure 31). A variation in thrust throughout the burn can be arranged by varying the cross-section along the length of the propellant grain.

Although the performance of solid propellants, in terms of thrust derived per unit mass, is generally inferior to that of liquid propellants, solids continue to be used in many applications. This is mainly because they are easier to handle and store and are simpler, therefore more reliable, in their mode of operation.

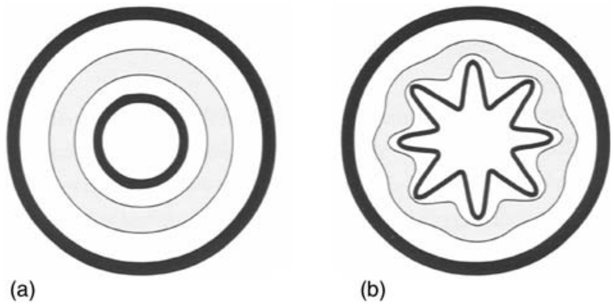


Figure 31. Solid-propellant grain cross-sections: (a) thrust increases throughout the burn; (b) constant thrust. [*Reproduced with permission from the Dictionary of Space Technology. Adam Hilger, 1900.*]

See also **Rocket Propulsion, Liquid Propellant; Space Launch Vehicles; Space Shuttle**

MARK WILLIAMSON

Further Reading

Brown, C.D. *Spacecraft Propulsion*. American Institute of Aeronautics and Astronautics, Washington D.C., 1996.

Davenas, A., Ed. *Solid Rocket Propulsion Technology*. Pergamon Press, Oxford, 1993.

Sutton, G.P. *Rocket Propulsion Elements: An Introduction to the Engineering of Rockets*. Wiley, New York, 1956.

Stoiko, M. *Pioneers of Rocketry*. Hawthorn Books, New York, 1974.

Turchi, P.J., Ed. *Propulsion Techniques: Action and Reaction*. American Institute of Aeronautics and Astronautics, Reston, 1998.

Williams, B. and Epstein, S. *The Rocket Pioneers*. Lutterworth Press, London, 1957.

Williamson, M. Spacecraft propulsion systems. *Phys. Technol.*, 15, 6, 284–290, 1984.

Winter, F.H. *The Golden Age of Rocketry: Congreve and Hale Rockets of the Nineteenth Century*. Smithsonian Institution Press, Washington D.C., 1990.

S

Satellites, Communications

Arthur C. Clarke, a science-fiction writer and an early member of the British Interplanetary Society, is credited as first stating the theoretical possibility of satellite communications. The February 1945 issue of the British technical journal *Wireless World* included a letter from Clarke under the title "V-2 for Ionospheric Research" in which he explained that three artificial satellites positioned 120 degrees apart in geosynchronous orbit could provide television and microwave coverage to the entire planet. Three months later, he privately circulated six typewritten copies of a paper titled "The Space Station: Its Radio Applications." A refined version, where Clarke gave a detailed, technical analysis of the orbital geometry and communications links, appeared under the title "Extra-Terrestrial Relays" in the October 1945 issue of *Wireless World.* Clark, however, acknowledged that the forerunner of the concept came from Hermann Potocnik, whose 1929 book *Das Problem der Befahrung des Weltraums* [*The Problem of Space Travel*], published under the pseudonym Noordung, described "stationary circling."

During the next decade, others promoted the idea of satellite communications. In their May 1946 report titled "Preliminary Design of an Experimental World-Circling Spaceship," Project RAND engineers at the Douglas Aircraft Company's plant in Santa Monica, California, told the U.S. Air Force that satellites could significantly improve the reliability of long-range communications and might spawn a multibillion-dollar commercial market. Subsequent RAND studies by James E. Lipp in February 1947 and Richard S. Wehner in July 1949 further developed the concept of equatorial-orbiting communications satellites. Writing under a pseudonym in *Amazing Science Fiction*, John R. Pierce of AT&T's Bell Telephone Laboratories suggested a communications satellite system in 1952. He became one of the first people outside defense-related circles to evaluate systematically both technical options and financial prospects. In a 1954 speech and a 1955 article, Pierce assessed the utility of passive "reflector" and active "repeater" satellites at various orbital altitudes and estimated a single satellite's capacity for handling simultaneous telephone calls as nearly 30 times greater than the original transatlantic telephone cable.

The USSR's launch in 1957 of Sputnik-1 (Figure 1), which transmitted an electronic signal back to earth simply for tracking purposes, sparked serious efforts by the U.S. to develop satellite communications for both military and commercial use. Score, developed by the Advanced Research Projects Agency and launched in December 1958, became the world's first active communications satellite. It received messages transmitted from a ground station and stored them on a tape recorder for retransmission back to earth. Launched in August 1960, the National Aeronautics American and Space Administration's (NASA's) Echo-1 first tested the merits of using a passive "reflector" satellite—a 30-meter-diameter, aluminized mylar balloon—for the transmission of voice, data, and photographs between ground terminals. The U.S. Army's Courier satellite, launched in October 1960, operated on much the same principles as Score but carried solar cells and rechargeable batteries to extend its potential lifetime. Stemming from its interest in transoceanic

Figure 1. The Sputnik-1 satellite on a rigging truck in the assembly shop in the fall of 1957 as a technician puts finishing touches on it.
[*Credit: Courtesy of NASA*.]

communication, AT&T launched Telstar-1 in June 1962 to experiment with telegraph, facsimile, television, and multichannel telephone transmissions between the U.S. and Europe, as well as Japan. With the launch of NASA's Syncom-2 in July 1963, the world finally had its first geosynchronous communications satellite. Not until 1975 would the USSR achieve a similar feat with its Raduga spacecraft for military and governmental communications, followed by its Ekran series in 1976 for direct television broadcasting and its Gorizont series in 1979 for domestic and international telecommunications.

Throughout the remainder of the twentieth century, most communications satellites were placed in geosynchronous orbit, because that afforded the greatest coverage with the smallest constellation. Established in 1964, the International Telecommunications Satellite Consortium (Intelsat)—with Communications Satellite Corporation (Comsat) as its U.S. component—opted for a geosynchronous satellite produced by Hughes Aircraft Company instead of a medium-orbit model proposed jointly by AT&T and RCA. Early Bird (Intelsat-1), launched on 6 April 1965, marked the beginning of global satellite communications networks open to all nations. Government-controlled, regional satellites to service a group of geographically proximate or culturally similar nations soon appeared—one of the first being Indonesia's Palapa-A in 1976.

Privately owned satellites represented yet another type of international system—the first being Japan Communications Satellite's JCSat-1 in 1989. The first domestic communications satellite was Telsat Canada's Anik-A in 1972, followed two years later by Western Union's nearly identical Westar-1, the first U.S. domestic communications satellite. During the 1980s, the International Mobile Satellite Organization (INMARSAT) operated a mobile telecommunications network primarily for maritime users but began expanding voice and facsimile services to aircraft on international routes during the 1990s.

While the U.S. military has relied heavily on commercial satellite communications over the years, it has also developed and launched dedicated systems to satisfy unique national security requirements. On 16 June 1966, the U.S. Air Force launched the first seven of 26 Initial Defense Communications Satellite Program (IDCSP) satellites, which provided high-speed digital data links from Vietnam to Washington D.C. That success led to the geosynchronous Defense Satellite Communications System (DSCS) II constellation during the 1970s (Figure 2) and the even more improved, jam-resistant, secure DSCS III in the 1980s (Figure 3). In December 1990, DSCS III links became the primary means for transmitting missile-warning data from key sensors worldwide to processing and command centers in the central U.S. To ensure survival of communications links, even during nuclear conflicts, the U.S. Air Force launched the first satellite in the Milstar program on 7 February 1994. A Fleet Satellite Communications (FLTSATCOM) constellation, provided links for the U.S. Navy from the late 1970s until replaced by ultrahigh frequency follow-on (UFO) satellites in the 1990s. The US also provided assistance in the development of Britain's Skynet military satellite communications system as well as a dedicated North Atlantic Treaty Organization (NATO) satellite communications system.

A competitive rush during the 1990s to provide global, commercial satellite telecommunications services to individual and corporate mobile phone users led to constellations of numerous, small satellites in low or intermediate (medium), rather than geosynchronous, orbit. Testing for Orbcomm's 26-satellite, data-transfer network began with the launch of an experimental package in 1991. Teledesic envisioned an 840-satellite design in 1994, but it was scaled back to 288 in 1998 and further reduced to only 30 in medium-earth orbit (MEO) in February 2002. Motorola's Iridium

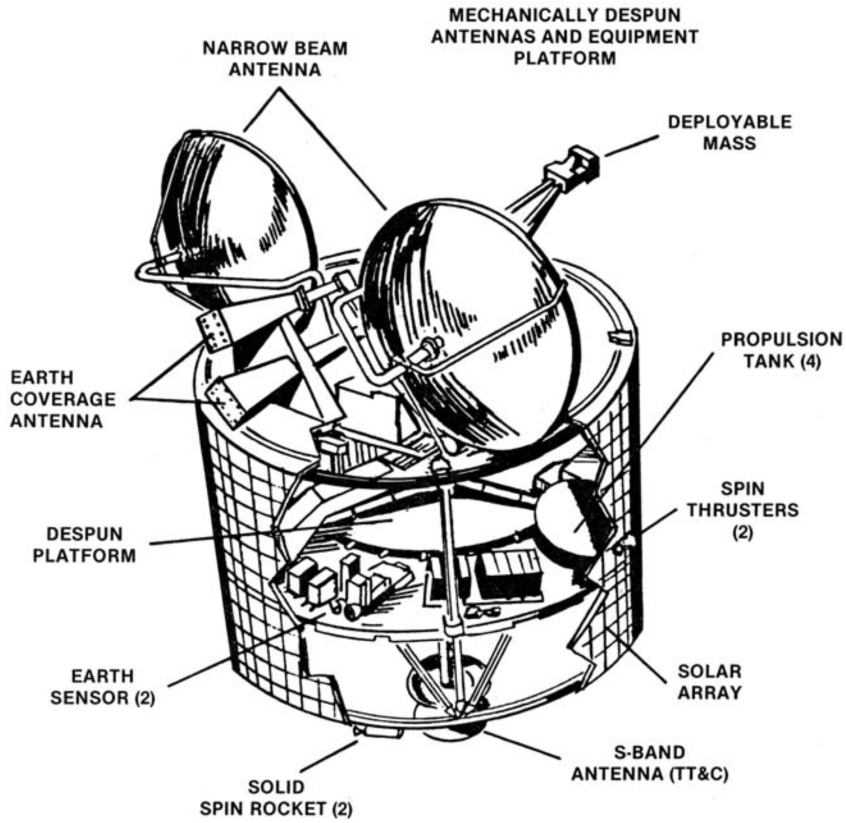


Figure 2. DSCS-II satellite diagram. [*Courtesy U.S. Air Force.*]

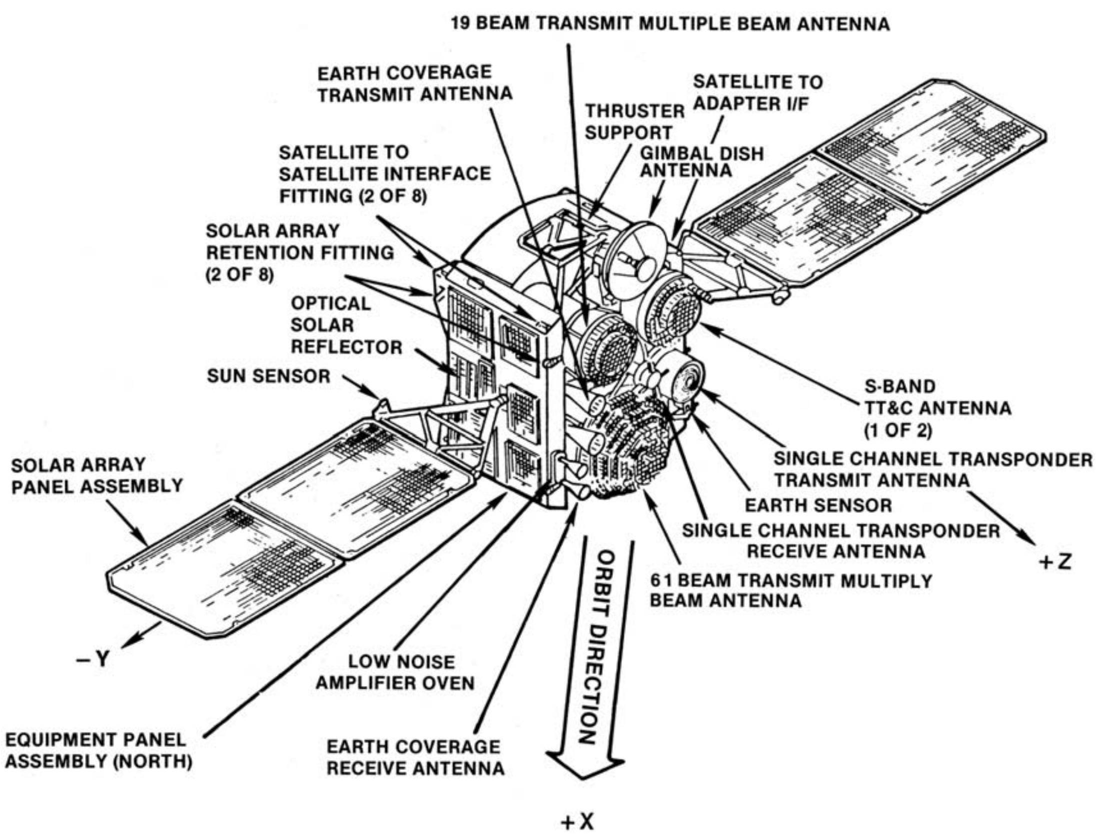


Figure 3. DSCS-III satellite diagram.
[*Courtesy U.S. Air Force.*]

system, which began with the launch of five satellites in May 1997, relied on 66 identical satellites in circular, low-earth orbit (LEO). Globalstar, which envisioned a 64-satellite constellation, entered the arena by launching four in February 1998. When customers failed to purchase services at the pace originally projected, a series of bankruptcies, buyouts, and reorganizations occurred. Nevertheless, prospects for ultimate success remained bright at the dawn of the twenty-first century.

See also **Sputniks**

RICK W. STURDEVANT

Further Reading

Clarke, A.C. *Ascent to Orbit: A Scientific Autobiography*. Wiley, New York, 1984. Contains all of Clarke's original technical writings, together with historical annotations in Clarke's own words.

Dorsey, G. *Silicon Sky: How One Small Start-Up Went Over the Top to Beat the Big Boys Into Satellite Heaven*. Perseus, Cambridge, 1999.

Hudson, H.E. *Communication Satellites Their Development and Impact*. Free Press, New York, 1990.

Johnson, N.L. and Rodvold, D.M. *Europe & Asia in Space, 1993–1994*. Kaman Sciences Corporation, Colorado Springs, ND.

Martin, D.H. *Communication Satellites*, 4th edn. Aerospace Corporation, El Segundo, 2000.

Spires, D.N. and Sturdevant, R.W. From ADVENT to Milstar: the United States Air Force and the challenge of military satellite communications. *J. Br. Interplanet. Soc.*, 50, 207–214, 1997.

Useful websites

National Aeronautics and Space Administration (NASA). *Communications Satellite History*: http://www.hq.nasa.gov/office/pao/History/commsat.html

Wood, L. *Lloyd's Satellite Constellations*: http://www.ee.surrey.ac.uk/Personal/L.Wood/constellations/index.html

Satellites, Environmental Sensing

Aerial photography was the precursor of satellite imagery for recording earth-surface characteristics using instrumentation that is located at a distance from (i.e., remotely), rather than in contact with, the subject. Established in the late 1800s, the first aerial photographs were obtained from cameras aboard balloons and kites until the advent of aircraft in the early 1900s. These early aircraft made a substantial impact on the collection of land-use and land-cover information, resource inventory, and archaeological data as well as being an important tool in wartime. In the late 1940s the first attempts at space exploration provided new opportunities for remote sensing of the earth. The first satellite for remote sensing was TIROS-1 launched in 1960; it was designed to collect meteorological data such as cloud cover. As improvements in sensors occurred, the range and quality of the data collected by meteorological satellites also improved. In particular the physics of the sensors became sufficiently sophisticated to compensate for distortions due to turbulence and refraction of light in regions of varying densities and temperatures in the atmosphere.

> *"The prospect of looking through,* not just *at,* the earth's atmosphere had begun." [*Lillesand and Keifer, 2000, p. 48*]

Further developments occurred as space flights increased. In the 1960s astronauts used hand-held cameras to capture images of the earth. Experiments involving color-reversal film to determine geological features were conducted successfully as part of the Gemini program. Subsequently, other aspects of the earth's surface became the focus of photographic record; for example, the oceans and the world's vegetation cover. New frontiers were breached with the Apollo space program, especially the use of mechanically operated Hasselblad cameras in conjunction with multispectral photography. The latter involves different types of film with various filters; for example, panchromatic (multicolored) film, infrared monochrome (black and white) film, and infrared color film. Piloted and robotic spacecraft with radar on board began as part of Russia's military intelligence program, but by the late 1970s weather radar (looking for radar echoes of precipitation) and radar altimeters (measuring heights of the earth's surface from the time delay of a signal) were deployed from satellites.

Radar sensor systems are "active" in that they emit the same microwave radiation that is used for the remote sensing. "Passive" sensors, on the other hand, are dependent upon receiving reflected sunlight or thermal infrared emission. All passive sensors record electromagnetic radiation (i.e., radiant energy) which is emitted from the earth's surface or atmosphere and which is classified according to wavelength. These types of radiant energy, listed in order of increasing wavelength, are gamma radiation, x-rays, ultraviolet radiation, visible radiation, microwaves, and radiowaves. As instrumentation have become increasingly sophisticated, sensors that can record a variety of radiant-energy wavelengths, (i.e., multispectral) have been developed. Skylab, launched in 1973 and carrying the earth resources experiment

package (EREP) was an early example of space-based multispectral remote sensing involving photographic and electronic measurements. From this the National Aeronautics and Space Administration (NASA) began to develop a program of recording earth resources from satellites. The first satellite in what later became the Landsat programme was launched in 1972. Landsat has become one of the most enduring earth observation missions, the latest satellite being Landsat 7 launched in 1999. The instruments on the Landsat satellites record data in the electromagnetic range between visible light and thermal radiation and with resolutions between 80 and 15 meters. The detail revealed by these missions confirmed the value of satellite remote sensing for monitoring the status of earth-surface features and resources. The Landsat program was the start of true satellite remote sensing. It shed light on a range of features, contributed to the explanation of their structure and formation and highlighted the magnitude of environmental change in many regions, including the rapidly declining extent of tropical forests.

Subsequent satellites for remote sensing include "Seasat," launched in 1978. Carrying radar sensors, it focused on the oceans, their circulation and sea-ice cover. The French *Centre National d'Etudes* (CNES*)* in combination with Belgium and Sweden launched the first non-U.S. satellite in 1986. This was one of a series known as *Systeme Pour l'Observation de la Terre* (i.e., SPOT), which subsequently developed into an international project involving five satellites launched between 1986 and 2002. Soviet satellite monitoring facilities included the Cosmos-1870 and ALMAZ-1 in 1987 and 1991 respectively. Both collected radar images and ALMAZ-1 was the first satellite to operate commercially. The European Space Agency (ESA) also launched a satellite, ERS-1, in 1991, and ERS-2 in 1995. The primary sensors record microwave, radar and radiometric data. Other nations that have developed satellite monitoring systems include Japan whose space agency launched JERS-1 in 1992, Canada launched its first Radarsat in 1995 and China launched the satellites YZ-1 and YZ-2 in 1999 and 2000. There were at least 31 satellites in orbit in 2000.

These satellites transmit images to earth and sophisticated computer programs translate data from various wavelengths into information on geology, water resources, vegetation, soils, minerals, tectonics, agriculture, forestry and urban environments. The end products are images from which maps of particular characteristics can be constructed. The continuous recording of earth-surface features provides detailed information on global environmental change at various scales. Such data can facilitate management, the prediction of future change and the assessment of the quality and quantity of biological and mineral resources.

The varied applications of satellite imagery have turned what began as government-funded endeavor into a commercial activity. However, satellites have also been developed for other purposes, two of which include weather and climate, and military applications. The recording of weather and climatic data involves monitoring of the atmosphere; today's efforts have advanced considerably on those of TIROS-1 (see above), and now weather and climate prediction, including advance warning of extreme conditions, is a primary objective. Satellites so designed are known as metsats; examples include those of the U.S. National Oceanic and Atmospheric Administration (NOAA), the Geostationary Operational Environmental Satellite (GOES), and the U.S. Air Force Defense Meteorological Satellite Program (DMSP). Military satellites differ from nonmilitary satellites insofar as they are always government funded, provide encrypted data that requires a special and secure receiver, and operate at a higher resolution.

Recent developments include a new family of agile satellites, such as IKONOS (1-meter resolution and Quickbird (0.61-meter resolution), which exploit emerging optical technologies to compete in the market for high-resolution images normally provided by aerial photography. The rapid and continuous collection of varied environmental data, the increasing resolution possible by sensors and increasing sophistication of analytical techniques mean that satellite monitoring has a bright future.

See also **Environmental Monitoring**

A. M. MANNION

Further Reading

Campbell, J.B. *Introduction to Remote Sensing*, 3rd edn. Taylor & Francis, London.

Dury, S.A. *Images of the Earth: A Guide to Remote Sensing*, 2nd edn. Oxford Science Publications, Oxford, 1998.

Konecny, G. *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*. Taylor & Francis, London, 2003.

Lillesand, T.M. and Keifer, R.W. *Remote Sensing and Image Interpretation*, 4th edn. Wiley, New York, 2000.

Monmonier, M.S. *Spying with Maps: Surveillance Technologies and the Future of Privacy*. University of Chicago Press, Chicago, 2002.

Useful Websites

Canada Center for Remote Sensing. *Learning Resources*: http://www.ccrs.nrcan.gc.ca/ccrs/learn/learn_e.html

Short, N.M. *The Remote Sensing Tutorial*: http://rst.gsfc.nasa.gov/start.html

Semiconductors, Compound

Compound semiconductors are semiconductors made up of two or more elements. Since the 1960s, group III and group V compound semiconductors (sometimes called intermetallic semiconductors) such as gallium arsenide (GaAs), gallium nitride (GaN), and indium phosphate (InP) have been important in semiconductor and optoelectronic devices such as lasers (fiber-optic transmission, CD players), light emitting diodes, and microwave integrated circuits (cellular phones), but early compound semiconductors featured in turn of the century devices are considered here first.

Experimental work on compound semiconductors began as early as 1874 when Karl Ferdinand Braun (Germany) described the rectifying properties of lead sulfide and iron sulfide crystals. The invention of radio communication acted as a stimulus to the development of the crystal detector, but due to the amplifying properties of the thermionic valve, interest in crystal detectors diminished during most of the interwar period. During World War I, lead sulfide and selenium photodetectors were developed in Germany capable of operating in the infrared spectrum, although conversion efficiencies were low (less than 1 per cent). Furthermore, selenium suffered from a slow response to changes in light, rendering it unsuitable for many applications. Work on lead sulfide detectors was resumed in Germany in 1932, where lead selenide and lead telluride were also investigated. In order to increase conversion efficiency, solid carbon dioxide and liquid nitrogen were used as coolants. Shortly afterwards, work on lead sulfide detectors began in Poland. An important commercial development was the introduction of the copper oxide rectifier (1927) and selenium rectifier (1931), both having advantages over thermionic diodes as low-frequency power rectifiers. Although manufactured in large quantities, no adequate theory existed at the time to explain their action. An important handicap affecting the manufacture of these devices prior to the early 1950s was the lack of suitable techniques of material purification. During the 1940s and 1950s, attention shifted towards germanium and silicon, initially in response to wartime demands for efficient high-frequency detectors for use in radar equipment.

The semiconductor industry has been largely dominated by the element silicon (Si) since the development of the planar process of device manufacture by Jean Hoerni of the Fairchild Corporation in 1959. However, group III–V compound semiconductors in particular have been increasingly employed from the 1960s onward in a growing number of applications. Some group III–V compound semiconductors have important advantages over silicon where high-frequency response and high-speed switching are important; they are also able to withstand higher temperatures and radiation damage. Silicon, because of its limited frequency response, is also restricted in its use as a detector of electromagnetic radiation in both the visible and infrared spectrum, and compound semiconductor materials are much more widely used in this field. Given these advantages, it is not surprising that military requirements in particular have stimulated research into the properties of compound semiconductors and their applications.

Interest in gallium arsenide (GaAs) began in the early 1950s as part of a search for higher band-gap materials that could operate at higher frequencies than silicon, in order to manufacture devices to meet military requirements for microwave communications and radars. Gallium arsenide was the first group III–V semiconductor compound to find widespread commercial application. It possesses the advantages mentioned above but, unlike silicon and germanium, suffers the disadvantage that (in common with other III–V semiconductor compounds) it exhibits a very low vapor pressure at its melting temperature. This causes additional problems in material preparation. Also, most compound semiconductor materials evaporate at different rates, altering their proportions during crystal growth. To prevent this, the required material is grown under pressure or prevented from evaporating by means of a nonreactive covering. Further disadvantages are that GaAs is difficult to dope with the required impurities, does not grow an electrically stable oxide, and crystal defects tend to be higher than those for silicon. Consequently, GaAs, like other III–V semiconductor compounds, is more difficult to manufacture than silicon, has taken more time to develop, and is therefore more expensive; this factor in particular so far preventing its wider use.

One characteristic of GaAs is that electron carrier mobility (and therefore electron transit time) is higher than in silicon. This fact can be

put to advantage in the manufacture of *n*-channel field-effect transistors, since channel conduction is entirely by electrons. Such devices may be three to five times faster than their Si equivalents. However, since hole mobility has a lower value in GaAs than in Si, *p*-channel GaAs field-effect transistors do not have this advantage.

GaAs has a band-gap of 1.43 electron volts. Devices have recently been developed made from wider band-gap compounds such as silicon carbide (SiC) and gallium nitride (GaN) (3.26 and 3.39 electron volts, respectively). They can therefore withstand a much wider range of temperatures ranging from the heat of a jet engine to conditions met with in outer space. Other advantages are high-voltage breakdown and low noise levels. A major application for such materials will be in the construction of high-power, high-frequency devices, although at present fabrication problems still remain.

Electrical conductivity may also be achieved by means of light. This phenomenon is known as photoconductivity, and was discovered in selenium by Willoughby Smith in 1873. The absorption of photons arriving on the surface of a suitable material may possess sufficient energy to raise electrons across the forbidden gap to the conduction band. Since different materials have different band gaps, they vary in response to photons of different wavelengths, and therefore vary from each other both in efficiency and spectral response. For example, Si and GaAs have peak spectral responses in the infrared (1100 and 870 nanometers, respectively), well outside the visible range (approximately 390 to 770 nanometers) Widely used compounds peaking near or within the visible spectrum are GaP (550 nanometers), GaN (370 nanometers), and SiC (470 nanometers).

Another important use of III–V compounds has been in the fabrication of light emitting diodes (LEDs), since silicon is very poor at emitting light. It was known by about 1960 that semiconductor materials could emit light. GaAs and GaAsP (with phosphorus) pulse-operated single *p–n* junction devices were made by Nick Holonyak in 1962, emitting red light. These materials were succeeded during the subsequent decade by indium arsenide (InAs), indium phosphate (InP), and indium antimonide (InSb). More complex structures then followed; for example, InGaAs, InAsP and InGaP. A significant advance (giving a two-fold increase in spectral efficiency) was made in 1990 when AlGaInP LEDs were constructed. By the end of the twentieth century they remained the most efficient LEDs within the range 570 to 650 nanometers (red–orange–yellow), achieving quantum efficiencies of up to 24 percent at 635.6 nanometers. These devices are claimed to be more efficient than unfiltered incandescent lamps, with projected lifetimes over a magnitude greater than conventional light sources.

By the mid-1990s, LEDs were being used over a color range from red to green. However, an unfilled gap existed in the blue region of the spectrum. In 1996, a breakthrough came with the production of high quantum-efficiency GaN devices and the blue laser diode by Shuji Nakamura at Nichia Chemical Industries in Japan. Previously, the best that could be achieved was by using SiC diodes, giving a peak quantum efficiency of 0.02 percent. Using GaN, quantum efficiencies of up to 10 percent in the blue region and 5 percent in the green was now possible, the latter figure also being a considerable improvement over other types.

In order to improve their characteristics, more complex multijunction devices have evolved from the simple *p–n* junction. Increased quantum efficiency and a wider range of spectral coverage has led LEDs to being used in an increasing number of applications in, for example, advertising displays, traffic signals, automobile stop lights, and indicators.

A further important use for compound semiconductor materials is as a laser light source, since stimulated emission at a definite wavelength may be produced by an applied voltage. Coherent radiation from a forward biased GaAs *p–n* junction was first observed in the U.S. (Gunter Fenner, Robert Hall, and Jack Kingsley at General Electric, 1962). The first diode laser optical recording system was produced in Holland by the Philips Company in 1978. It used an AlGaAs laser capable of delivering a pulsed light output equivalent to a large gas laser, despite its small size. It has since (as the CD player) become the major domestic application for lasers.

Another widely growing use for lasers is as a light source in the field of optical fiber transmission. In this application, the device must be optically matched to the transmission line in order to achieve maximum efficiency. To achieve this, advantage is taken of the property of group III–V compounds to undergo variation in peak spectral response when doped with minute amounts of further components whilst in the molten state. For example, phosphorous is used in this manner to dope GaAs. The ternary mixture thus formed is GaAsP. A further example is the widely used material AlGaMP.

Following work in the U.K., it was demonstrated in the U.S. (by John Gunn, 1963) that when an electrical field is applied to certain group III–V semiconductors, microwave oscillations take place when the field exceeds a certain value. Such a device is known as the Gunn diode oscillator. III–IV compounds have also found applications in the field of infrared technology. In 1960, E.H. Putley, of the U.K. Royal Radar Establishment, observed photoconductivity at low temperatures in *n*-type InSb at wavelengths between 0.1 and 4.0 millimeters. Subsequent work in this field has been largely in connection with military applications.

See also **Lasers, Applications; Lasers in Optoelectronics; Light Emitting Diodes; Semiconductors, Elemental; Semiconductors, Postband Theory; Semiconductors, Preband Theory**

ROBIN MORRIS

Further Reading

Bar-Chaim, X., Ury, I. *et al.* Integrated optoelectronics. *IEEE Spectrum*, 5, 38, 1982.

Bell, T.E. Japan reaches beyond silicon. *IEEE Spectrum*, 10, 46–52, 1985.

Braunstein, R. Radiative transitions in semiconductors. *Phys. Rev.*, 99, 195, 1892.

Denbaars, S.P. Gallium-nitride-based materials for blue to ultraviolet optoelectronics devices. *IEEE Proc.*, 85, 11, 1740–1749, 1997.

Dettmar, R. GaAs on silicon—the ultimate wafer? *ME Review*, 4, 136–37, 1989.

Kenney, G.C., Lou, D.Y.K. *et al.* An optical disc replaces 25 mag tapes. *IEEE Spectrum*, 2, 33, 1979.

Loebner, E.E. Subhistories of the light emitting diode. *IEEE Trans.*, ED-23, 675–698, 1976.

Martin, R. In search of the blue light. *Chem. Ind.*, 5, 173–177, 1999.

Nakamura, S. and Fasol, G. *GaN Based Light Emitters and Lasers*. Springer, New York, 1997.

Vanderwater, D.A., Tan, I.H. *et al.* High brightness AlGaInP light emitting diodes. *IEEE Proc.*, 85, 11, 1752–1764, 1997.

Voelcker, L. The Gunn Effect. *IEEE Spectrum*, 7, 24, 1989.

Semiconductors, Crystal Growing and Purification

Crystals of high purity are essential to the manufacture of semiconductor devices, including integrated circuits, and much effort has gone into their refinement. The first single semiconductor crystal was drawn in 1948 and, from 1949 or 1950 onward, monocrystalline material has been used exclusively. The conditions imposed are extremely stringent and have become progressively more so as manufacturing technology has advanced. From the time of the invention of the transistor in 1947 until the mid-1960s most were made using germanium, although since then silicon has been mainly employed. By the 1980s silicon accounted for 98 percent of all semiconductor devices sold worldwide.

The three basic stages in the production process are extraction, purification, and crystal growing.

Extraction

The extraction of metallurgical grade silicon takes place within an electrode-arc furnace. Quartzite, coal, coke, and wood chips are fed into the furnace and a series of chemical reactions take place, finally resulting in the extraction of silicon of about 98 percent purity. The major contaminant is usually boron, although carbon and other impurities are also present. Only a small fraction (about 1 percent) of silicon produced by these means is used in the semiconductor industry.

Purification

This stage is necessary in order to produce electronic grade silicon, a material polycrystalline in structure with impurity levels of only a few parts per billion. Such low impurity levels are necessary to ensure that contamination does not occur during the following crystal-growing stage. The purification process currently used is basically as follows:

1. Pulverised metallurgical grade silicon is heated with anhydrous hydrogen chloride at 300°C, yielding trichlorosilane (a gas), together with hydrogen.
2. The trichlorosilane is cooled and becomes liquid (at 32°C).
3. Fractional distillation is employed to remove impurities present (especially phosphorous and boron).
4. The reaction is reversed by passing a mixture of trichlorosilane and hydrogen over a resistance-heated silicon rod at a temperature of between 1000 and 1200°C.
5. Small crystals of electronic grade silicon are deposited on the rod during the reaction. Pure polycrystalline rods so formed are a few meters long and several inches in diameter.

Due to the complexity and cost of the technology, it had become limited to the U.S., Japan and West Germany well before the end of the twentieth century.

An earlier method of material purification, zone refining, was first described by the Russian physicist Petr Kapitza in 1928 and perfected at Bell Laboratories by William G. Pfann in 1951. The impure silicon or germanium polycrystalline rod is supported in a quartz tube in an inert ambient, and a small zone melted by radio-frequency heating. The impurities then collect at the zone edge where the temperature of the melt is lowest. When the zone is gradually swept from one end of the rod to the other the impurities are swept along the rod and collected at the end. This effect occurs because most impurities have a lower melting point than silicon or germanium, which therefore crystallize out of the melt before the impurities. Although a great advance upon earlier methods, zone refining has the disadvantage that only relatively small amounts of material can be purified and the removal of boron in particular is time consuming.

Crystal Growing

First, it is necessary to ensure that the material is as dislocation-free as possible, since crystal defects affect the rate at which impurities diffuse. Dopant atoms (introduced in highly accurate amounts during subsequent production processes) travel much more quickly at the surface of grain boundaries due to crystal dislocations. Dopant atoms introduced into material with an unacceptable level of dislocations would result in a highly uneven diffusion profile, and consequent device failure. It is also important to control the crystal orientation. This is because: (1) crystals cleave easily in certain directions; (2) during device manufacture, it is invariably necessary to put down uniform layers of dopants and oxides; and (3) chemical etching is frequently used to cut windows in oxides grown on the surface of wafers. Etching will only proceed uniformly if the crystal surface presents the correct orientation.

The favored crystal-growing technique is the Czochralski process, developed by Jan Czochralski in 1918, but perfected for growing germanium and silicon single crystals by Gordon K. Teal and John E. Little at Bell Laboratories in 1950. By the end of the twentieth century it accounted for 70 to 80 percent of production. Three other methods of crystal growing are also used, float-zone, Bridgman, and epitaxy. Each has its advantages and disadvantages.

In the case of the Czochralski method, the crystal is "pulled" from molten material (melt), which is contained within a crucible. There is therefore a possibility of material contamination. With germanium (melting point 960°C) it is possible to use a graphite crucible. However, this cannot be done with silicon (with a melting point of 1420°C) and therefore quartz is used. The process takes place within an enclosed inert ambient (helium or argon). The temperature of the melt is controlled externally by radio-frequency heating. The seed crystal, mounted in the desired orientation, is held in a chuck fixed at the end of a shaft, which is driven by a motor mounted externally above the enclosure. The crystal is then lowered to make contact with the melt, which is set at about 15°C above its melting point (about 1435°C). The crystal itself will not melt because of its higher resistivity. The seed crystal is then slowly raised and rotated, causing the molten silicon to freeze onto the seed crystal with the same orientation as the seed itself. Finally, the fully grown crystal is cooled to about 300°C before being exposed to the external atmosphere.

The float-zone method developed by P.H. Keck in 1953 overcomes the difficulty of crucible contamination by vertically suspending a polycrystalline silicon rod in an inert ambient. A radio-frequency coil melts a narrow cross-section of the rod above the seed crystal and is slowly raised. The silicon below the molten zone resolidifies in single-crystal form with the same orientation as the seed crystal. The molten zone does not flow out, due to the geometry and surface tension. This process is more expensive and difficult to control than the Czochralski method.

The Bridgman method (named after Percy W. Bridgman) involves a boat containing seed and melt being slowly pulled through a horizontal tube furnace, the melt freezing in the required orientation as the crystal enters the cooler zone. Towards the latter years of the twentieth century this was still the preferred technique for growing gallium arsenide (GaAs) crystals. Single crystals of 75 millimeters diameter and low dislocation density were by then being produced on a production basis.

Conclusion

All the above processes involve crystal growth from the melt. Epitaxy is the technique of depositing single-crystal material in successive layers upon an atomically flat single crystal substrate. The substrate most commonly used is silicon, which enables many compound semiconductors to be deposited; for example, crystalline GaAs and cadmium telluride (CdTe). During the epitaxial

process the deposited material takes up the same crystalline orientation as the substrate. Fabricating transistors by impurity doping of epitaxial single crystals grown from the gas phase was first achieved at Bell Laboratories in 1960. In the following year, N. Nielson of RCA described the process of epitaxially growing GaAs and Ge crystals from the liquid phase. Molecular beam epitaxy, although originating in elementary form in the 1960s, only came into production use by the early 1980s. It is however capable of greater precision.

See also **Integrated Circuits; Semiconductors, Compound; Semiconductors, Elemental; Semiconductors, Postband Theory; Thin Film Materials and Technology**

ROBIN MORRIS

Further Reading

Gilman, J.J. *The Art and Science of Growing Crystals*. Wiley, New York, 1963.

Harper, C.A. Handbook of Materials and Processes for Electronics. McGraw-Hill, New York, 1970.

Keck, P.R. and Golay, M.I.E. Crystallisation of silicon from a floating liquid zone. *Phys. Rev*., 89, 1297, 1953.

Pearce, C.W. Crystal growth and wafer preparation, in *VLSI Technology*, Sze S.M., Ed. McGraw-Hill, Singapore, 1985, pp. 9–50.

Pearce, C.W. Epitaxy, in *VLSI Technology*, Sze S.M., Ed. McGraw-Hill, Singapore, 1985, 51–92.

Petritz, R.L. Contributions of materials technology to semiconductor devices. *IRE Proc*., 1025–1037, 1962.

Pfann, W.G. Techniques of zone melting and crystal growing, in *Solid State Physics*, Seitz, F. and Turnbull, D., Eds. Academic Press, New York, 1957, pp. 423–521.

Teal, G.K. and Buehler, E. Growth of single crystal silicon crystals and of single crystal silicon *p–n* junctions. *Phys. Rev*., 87, 190, 1952.

Teal, G.K. and Little, I.E. Growth of germanium single crystals. *Phys. Rev*., 78, 647, 1950.

Thomas, R.N., *et al*. Status of device-qualified GaAs substrate for GaAs integrated circuits. *IEEE Proc*. 76, 7, 778–791, 1988.

Semiconductors, Elemental

Elemental semiconductors may be classified either as intrinsic or extrinsic. An intrinsic semiconductor is one that exists in pure crystalline form, its conductivity depending entirely upon its intrinsic properties. An extrinsic semiconductor is one containing impurity atoms within its crystalline structure. The addition of minute quantities of impurities (dopants) greatly affects the electrical properties of semiconductors, and they are therefore described as being structure sensitive. Because semiconductor materials only naturally occur in an extrinsic state, their value has largely depended upon developing techniques of refinement in order to make efficient use of their electrical properties.

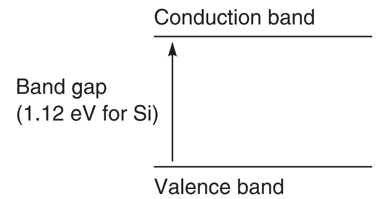


Figure 4. Electrons are raised from the valence band to the conduction band provided they have enough energy to jump the gap. Charge conduction is therefore directly related to band gap width.

An important factor in the selection of a semiconductor material is that its characteristics strongly depend upon the structure of the energy bands occupied by electrons. The band gap for a semiconductor is the energy gap between the valence (ground state) band and the conduction band (Figure 4). Materials with wide band gaps such as carbon in the form of diamond (band-gap energy 7 electron volts) may offer in theory the prospect of very high temperature operation since its wide band-gap means much higher energies are required to change the conducting state (other semiconductors change state at high voltage or high power). However diamond has so far remained virtually unused in electronic devices, due to practical difficulties in fabrication. Those semiconductors with narrow band gaps such as grey tin (0.1 electron volts) have also so far been found to be unsuitable. The principal elemental semiconductors used in device manufacture to date have been the group IV elements germanium (Gi) (0.72 electron volts) and silicon (Si) (1.12 electron volts), the material being initially purified to an intrinsic state and controlled amounts of dopants then added in order to meet the desired specifications.

The first successful application of elemental semiconductor materials was as point contact rectifiers for the detection of electromagnetic waves in the early twentieth century (see Radio Receivers, Crystal Detectors). From 1900 to 1910 a wide range of elements and compounds were investigated, lead sulfide (PbS) and silicon proving the most efficient. Although the characteristics of these devices were substantially improved over the following years, this was almost entirely as a result of empirical affect. Nevertheless, a variety of theories were put forward in an effort to explain the phenomenon of rectification. Early explanations included, for example, the thermoelectric

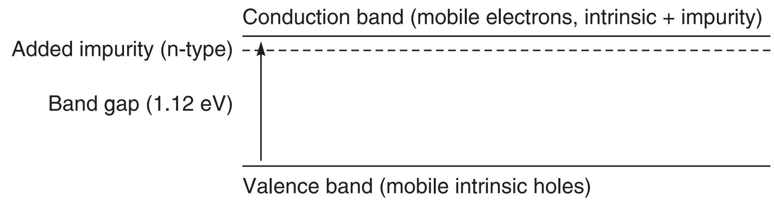


Figure 5. By adding minute controlled amounts of certain impurities to, for example, intrinsic silicon (group IV in the periodic table), the conductivity may be varied over a wide range. With the addition of *n*-type impurities (group V in the periodic table), electrons now become the majority carriers as the *n*-type impurity level donates electrons to the conduction band. Little energy is required due to the proximity of the impurity and conduction bands (0.05 electron-volt gap).

effect (already rejected by Karl Ferdinand Braun), the electrostatic barrier theory, propounded by M. Huizinga (Germany) in 1920, and the cold emission theory of G. Hoffman (Germany) in 1921. About this time, Walter Schottky (Germany) also put forward a semiconductor theory with blocking layers and potential thresholds, but did not yet realize that holes would act as charge carriers. However, none proved satisfactory prior to the band theory of electronic conduction in solids, developed by Alan H Wilson (U.K.) in 1931. By the beginning of World War II a theoretical picture was beginning to emerge that explained the movement of charge carriers in semiconductor junctions. B. Davydov (USSR) produced the first model of a *p–n* junction in 1938, which included the concept of minority carriers (holders) in the semiconductor conduction. This was followed by an explanation of matter and to semiconductor junctions by Shottky in 1939.

A major problem facing the early investigators was their lack of appreciation of the extremely high degree of purity required in order to control the equilibrium concentration of mobile charge carriers within semiconducting material, and hence keep its resistivity correct. Consequently, experimental results were unpredictable and difficult to evaluate. In any case, no suitable process of refining semiconductor material to the level of purity required was then available. A further important factor delaying progress was that, due to the overwhelming success of the thermionic valve, little stimulus existed until almost the end of the 1930s to carry out investigations into the properties of semiconductor materials. Gordon Teal (U.S.), recalling his doctoral thesis on germanium during that decade remarked, "Its complete uselessness fascinated and challenged me."

The urgent need for efficient ultrahigh frequency detectors for radar applications at the outbreak of World War II stimulated renewed interest in crystal detectors within the major advanced industrial countries. Isolated effort by individual researchers was now supplanted by the power of large research establishments and commercial corporations, with major emphasis being placed on the production of silicon and germanium point-contact diodes. In Britain, by 1940, the first commercially produced silicon diodes for use in radar frequencies were being made at Thomas Houston Limited, using polycrystalline silicon of 98 percent purity. Later, crystals were grown at the General Electric Company from highly purified silicon powder, to which had been added controlled fractions of aluminum and beryllium. Although great efforts were now being made in material preparation, the work largely proceeded on an ad hoc basis.

After 1941, the center of attention of semiconductor materials research shifted to America and

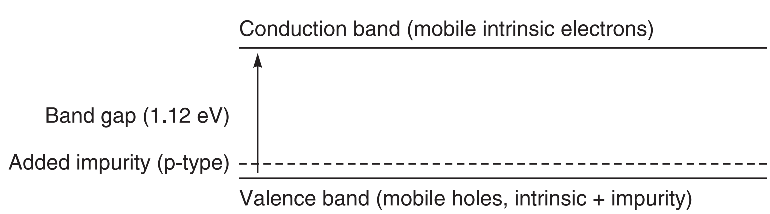


Figure 6. With the addition of *p*-type impurities (group III in the periodic table), holes now become the majority carriers as electrons are raised from the valence band to the *p*-type impurity band. Little energy is required due to the proximity of the impurity and valence bands (0.05 electron-volt gap).

to Bell Laboratories in particular. One early wartime method of material preparation was to produce pure germanium and silicon films by pyrolitic deposition on to tantalum filaments. It was then used to manufacture infrared detectors and photoconductors. The technique of material purification by the segregation of impurities through the repeated freezing was also employed at this time. From February 1942 onwards, germanium microwave diodes were being manufactured in quantity at various establishments and a basic research program on germanium was instituted. Pennsylvania University concentrated on silicon, and cooperating closely with industry, was soon producing material with a spectroscopic purity better than 99.9 percent using a process involving the reduction of silicon tetrachloride with zinc. Other firms, including Sperry, were also manufacturing microwave diodes.

Throughout the period of the war, polycrystalline material was used exclusively in the manufacture of semiconductor devices. However, in October 1948, Gordon Teal and John Little (Bell Laboratories) grew the first single germanium crystals by a crystal-pulling process developed initially by Jan Czochalski in 1918. Within months, crystals were being doped with impurities to produce germanium "grown junction" diodes. The period 1952–1953 was particularly important in terms of material purification and the consequent evaluation of the physical properties of semiconducting materials. The process of zone refining was introduced by Willian Pfann in 1952, and float-zone refining by Henry C Theurer in 1953. From then on, it was possible to produce material of a consistently high quality. In 1952 Teal and Ernest Buehler also produced large high-quality silicon crystals by pulling from the melt, which was impurity doped to form single crystal diodes. With such high purity material available, much more accurate measurements of electron and hole carrier mobility could now be carried out, consequently achieving a deeper understanding of the properties of intrinsic semiconductors.

In 1954, silicon grown-junction transistors, with a high degree of lattice orientation, were first fabricated by Gordon Teal at Texas Instruments, using a crystal-pulling process. The ability of silicon (due to its wider band gap) to operate at higher temperatures than germanium was of particular importance. Although methods of refining semiconductor materials have since been greatly improved, progress has been of a steady, incremental nature. The technique of growing ever-larger silicon crystals on a mass production basis, free from dislocations and other defects, has also constantly advanced. In 1975, crystal diameters were typically around 50 to 75 millimeters. By the end of the century, diameters had risen to 300 millimeters. From the mid-1960s onwards, silicon has almost entirely replaced germanium in device manufacture, a major factor having been the ability to grow an electrically stable oxide on its surface, thus rendering it suitable for planar fabrication.

The group VI element selenium (Se) has played an important role in the fabrication of electrical and electronic devices. From 1931 onwards, selenium rectifiers were used commercially on an increasing scale. The first practical solar cells, using selenium, were constructed as early as 1883. Their conversion efficiency was very low (less than 1 percent) and further development was neglected until the early 1930s, when they were introduced commercially. Cheap to produce, selenium photovoltaic cells were then employed in an increasingly wide range of applications, including photographic work. This use was aided by a peak spectral response (556 nanometers) approximating to that of the human eye. However, when used as photo detectors, conversion efficiencies still remained at about 1 percent and response time was limited. A significant breakthrough came with the development of the silicon photovoltaic cell by Russell Ohl (U.S.) in 1941. By 1944, silicon photovoltaic cells with conversion efficiencies of 6 percent were being manufactured, this figure rising to about 12 percent by 1960 and to about 16 percent by 1996. Silicon photocells have therefore replaced selenium, being more efficient, mechanically robust, and cheaper to manufacture.

The group IV element carbon (C) has also been widely used within the radio and electronics industry. Apart from possessing a negative temperature coefficient, it has the property of decreasing its resistance under applied pressure. This effect was utilized in the invention of the carbon microphone by Thomas A. Edison (U.S.) in 1877. A transverse-current carbon microphone, developed in Germany by G. Neumann in 1924, was used by the BBC between 1926 and 1935.

Carbon-film resistors were first made by T.E. Gambrell and A.F. Harris (U.K.) in 1897. The high-stability cracked-carbon type was invented at Siemens and Halske (1925) and the sprayed metal-film type a year later by S. Lowe (Germany). Other uses include automatic voltage regulators, electrodes, and brushes in the electrical machinery.

See also **Semiconductors, Compound; Semiconductors, Preband Theory, Semiconductors, Postband Theory**

ROBIN MORRIS

Further Reading

Brice, J.C. and Joyce, B.A. The technology of semiconductor materials preparation, in *Radio Electron. Eng.*, 43, January February, 21–28, 1973.
Fraser, D.A. *The Physics of Semiconductor Devices*. Clarendon Press, Oxford, 1977.
Fuller, C.S. and Ditzenburger, L.A. Diffusion of donor and acceptance elements in silicon. *J. Appl. Phys.*, 27, 544, 1956.
Pfann, W. Zone refining. *Scientific American*, 12 63–72, 1967.
Ryder, J.D. *Electronic Fundamentals and Applications*. Pitman, London, 1977.
Shaw, D.F. *An Introduction to Electronics*. Longman, London, 1962.
Shockley, W. *Electrons and Holes in Semiconductors*. Van Nostrand, NJ, 1963.
Teale, G.K. Single crystals of germanium and silicon—basic to the transistor and Integrated Circuit. *IEE Trans. Electron. Devices*, 23, 7, 621–639, 1976.

Semiconductors, Postband Theory

The theoretical explanation of transistor action rests upon the concepts of postclassical physics. Important advances took place from the early 1930s onward, when Arnold Sommerfeld, Felix Bloch and co-workers applied quantum theory concepts to the theory of metals. A study of the structure of semiconductor materials applying quantum mechanics was made in Britain by Alan Wilson, 1931–1932. However, only after World War II did clearer pictures begin to emerge which could explain the movement of charge carriers in semiconductor *p–n* junctions and metal-to-semiconductor contacts.

Semiconductor research was stimulated during the 1930s by the possibility of military applications in radar detectors and signaling. The major impetus came from the inability of thermionic valves, or vacuum tubes, to operate as fast switches, due to high interelectrode capacitance. This limited their use as signal detectors in radar equipment. Since the point contact diode proved superior in this respect, great efforts were made under wartime conditions to improve their characteristics. Consequently, large well-funded research programs were rapidly instituted by Britain and Germany, and also by the U.S. after 1941.

Although it was known that the addition of minute amounts of certain impurities to bulk semiconductor material altered the characteristics of semiconductor diodes, the effect was difficult to predict. This was because techniques of material purification were not sufficiently developed and therefore inconsistency in batch production was unavoidable. Consequently, work proceeded largely on an ad hoc basis. A method of obtaining a high degree of material purification (zone refining) was described by Petr Kapitza in 1928 but its significance was not realized until much later in the U.S. by William G. Pfann in 1952. The effects of various doping elements was still not clearly understood in the immediate postwar period. Petritz (U.S.) mentions that even as late as 1948 rectifiers were made with tin-doped germanium in the belief that tin was the doping element, although in fact doping levels were due to impurities within the tin, the tin itself having no electrical effect on the germanium.

Perhaps the most important semiconductor research during World War II was carried out at Purdue University and Bell Laboratories. Purdue concentrated on the study of germanium, Bell on improving silicon point-contact diodes. These devices had already been substantially improved in Britain and elsewhere since the early "cat's whisker" and now consisted of purified *p*- or *n*-doped crystalline material, the metal to semiconductor contact being made by a pointed wire (usually of tungsten or molybdenum) electrically attached by a "forming" process, resulting in a low capacitance rectifying *p–n* junction. The assembly was then sealed within an inert ambient.

The first *p–n* junction silicon rectifier was produced by Russell Ohl in 1941 and the presence of group III and V impurities in germanium and silicon was also discovered at about this time (Jack Scaff and William Pfann at Bell Labs). Consequently, it was now possible to make rectifying devices with much higher reverse breakdown voltages and handling much higher powers, as well as multijunction devices such as thyristors. Instead, efforts were concentrated on satisfying wartime requirements and by 1945 Bell (Western Electric) were producing over 50,000 rectifiers monthly.

The point-contact transistor (see Figure 7) was invented at Bell laboratories by Walter Brattain and John Bardeen, (1947), following an investigation of the surface properties of semiconductors and an understanding of the role of minority current carriers in electrical conduction. It consisted of two closely spaced wires in electrical contact with a substrate of *n*-type single-crystal germanium. This device was the product of a two-year goal-oriented research program. Realizing its significance, the U.S. government immediately

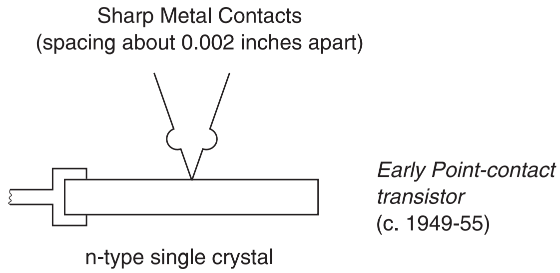


Figure 7. Early point-contact transistor, c. 1949–1955.

allocated funds for further development. Germanium was chosen because it had been investigated in detail during the war years; it was simpler in structure than compound semiconductors and its melting point (96°C) was lower than silicon (143°C).

Limited production of germanium point contact transistors began in 1951, but the device suffered from considerable defects. Its characteristics varied widely from one device to another, it was electrically unstable, and noise levels were high. Also, device action was extremely complex due to surface effects, making theoretical analysis very difficult. Efforts were therefore concentrated on the more electrically stable junction transistor.

The Czochralski crystal-growing process developed by Jan Czochralski in 1918 was a major advance allowing single crystals to be drawn from the molten state in the form of a *p–n–p* or *n–p–n* structure. This method enabled germanium junction transistors to be manufactured in quantity with uniform characteristics. An alternative approach, the germanium alloy-junction transistor (see Figure 8), had the advantage of superior frequency response. The U.S. military realized the significance of these improvements and immediately instituted a "million a month" manufacturing program with Western Electric.

The next major step was the successful manufacture of silicon grown-junction transistors by Gordon Teal in 1954. These were single crystal *n–p–n* junctions grown by the Czochralski process.

Again, this device was of great interest to the military, because it could operate at higher temperatures than germanium and from this time onwards large-scale military involvement took place within the industry, concentrating on silicon technology.

Two significant advances now followed: (1) precise control of junction depth by vapor diffusion; and (2) oxide masking during the diffusion process. Precise control of junction depth resulted in greatly extending high-frequency performance,

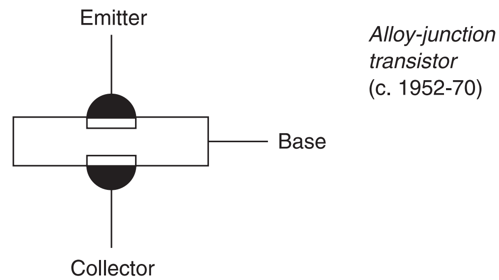


Figure 8. Alloy-junction transistor, c. 1952–1970. The germanium-alloy junction transistor has indium emitter and collector beads alloyed in a germanium base. Parameters are controlled by doping levels, geometry, and surface conditions.

while oxide masking electrically stabilized the device surface, greatly lowering leakage currents and improving voltage breakdown levels. These developments led to major advances in device construction, most significant by far being the planar transistor (see Figure 9). This device was so described because the insulating oxide covering the semiconductor surface formed a flat, planar layer. The planar approach offered increased reliability at decreased cost and also permitted large-scale integrated circuit manufacture. Apart from applications such as high-voltage rectifiers and thyristors, it has rendered previous methods of construction obsolete.

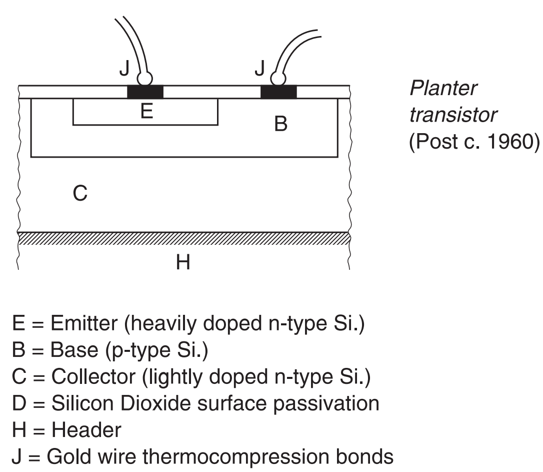


Figure 9. Planar transistor, post-1960. Note that the chip is welded to a gold-plated metallic substrate (header). Gold wires are bonded by thermocompression onto the aluminum bonding pads attached to emitter and base. The wires are then connected to electrically isolated posts on the header. The whole assembly is encapsulated in an inert ambient. Device parameters are controlled by geometry and doping levels.

In 1960, Bell Laboratories succeeded in growing very thin layers of doped semiconductor crystalline materials upon a silicon semiconductor substrate by vapor deposition. This process is termed epitaxial and it enabled high-resistivity layers to be deposited upon low-resistivity substrates, which cannot be done employing conventional diffusion techniques. It was now possible to construct devices within the epitaxial layer using advantageous doping profiles unattainable by previous means, this greatly assisting the subsequent development of integrated circuitry.

Planar technology is now in almost universal use and enables transistors, capacitors and resistors to be fabricated on a single slice with great precision. Windows are opened in the oxide by selective etching, enabling successive *p–n* junctions to be formed by impurity diffusion and also allowing metallic interconnections to be deposited. Integrated circuits use these metallic interconnections to link components electrically and achieve the desired circuit configuration. Since all the devices or integrated circuits on the slice are fabricated simultaneously, variation in their characteristics is minimal and only at the end of the production process are they separated to form "chips." Electrical connections from the chip to its external packaging are usually made using gold wires, which are attached to the bonding pads by thermocompression bonding. Development of integrated circuitry has largely rested upon silicon, since a stable germanium oxide cannot be grown successfully. However, a number of compound semiconductor materials were increasingly used.

Planar fabrication is particularly suited to the manufacture of metal-oxide field-effect transistors. This led to a whole new development in integrated circuitry from the late 1960s onward. Subsequently, component density per chip has approximately doubled every 18 months (Moore's law), resulting in great improvements in reliability, speed of operation, reduction in power dissipation and attainment of greater circuit complexity. Cost per bit has been vastly reduced. For example, a 1-kilobit dynamic random access memory (DRAM) manufactured in 1974 cost one cent per bit. By 1985, a 1-megabit DRAM was being produced as a cost of one thousandth of a cent per bit.

See also **Semiconductors, Compound; Semiconductors, Elemental; Semiconductors, Preband Theory, Integrated Circuits, Fabrication; Semiconductors: Crystal Growing, Purification; Transistors**

ROBIN MORRIS

Further Reading

Aschner, L.E., Bittman, C.A. *et al.* A double diffused silicon high frequency transistor produced by oxide masking techniques. *J. Electrochem. Soc.*, 106, 1145–1147, 1959.

Braun, E. and Macdonald, S. *Revolution in Miniature: The History and Impact of Semiconductor Electronics*. Cambridge University Press, Cambridge, 1978.

Frosch, C.J. and Derrick, L. Surface protection and selective masking during diffusion in silicon. *J. Electrochem. Soc.*, 104, 547, 1957.

Goldstein, A. and Aspray, W. Eds. *Facets: New Perspectives On the History of Semiconductors*. IEEE Center for the History of Electrical Engineering, New Brunswick, NJ, 1997.

Gorton, W.S. The genesis of the transistor. *IEEE Proc.*, January, 50–52, 1998.

Morris, P.R. *A History of the World Semiconductor Industry*. Peter Peregrinus, London, 1990.

Petritz, R.L. Contributions of materials technology to semiconductor devices. *IRE Proc.*, May 1025–1038, 1962.

Riordan, M. and Hoddeson, L. *Crystal Fire: the Birth of the Information Age*. W.W. Norton, London, 1997.

Roberts, D.H. Silicon integrated circuits. *Electronics and Power*, April, 282, 1984.

Ross, I.M. The invention of the transistor. *IEEE Proc.*, January, 1998.

Shockley, W. Twenty-five years of transistors. *Bell Labs. Record*, December 340–341, 1972.

Sparkes, J.J. The first decade of transistor development. *Radio Electron. Eng.*, 43, 8, 1973.

Semiconductors, Preband Theory

Investigations into the properties of what we now define as semiconductors began in the early nineteenth century. No satisfactory theoretical explanations were possible, and little effort was made to put discoveries to practical use. However, by the beginning of the twentieth century the main distinguishing characteristics of semiconductors (negative temperature coefficient of resistance, asymmetric conduction of electricity in solids, photoelectric effect, photovoltaic effect, large Hall effect, and high thermoelectric power) had been discovered. The one-way conductance, or rectification, property of semiconductors is the characteristic that makes semiconductor devices useful today.

Charles E. Fitts (U.S.) constructed the first practical photocell around 1883–1886 by spreading a semitransparent sheet of gold leaf onto the surface of a selenium layer on a copper backing plate. However, this early work was not pursued for some decades. The main stimulus leading to the development of semiconductor devices came through the need to develop an efficient detector of electromagnetic waves. Jagadis C. Bose (India) patented the first solid-state point-contact detector

in 1904, mentioning a variety of substances including galena (lead sulfide). In 1906 Henry Dunwoody and Greenleaf Whittier Pickard (U.S.) developed detectors using silicon carbide and silicon. In the first decade of the twentieth century many workers were engaged in efforts that led to improvements in detection efficiency for radio receivers (see Radio Receivers, Crystal Detectors and Receivers). This resulted in production of the so-called "cat's whisker," consisting of a rectifying metal to semiconductor contact, the polycrystalline surface of the material being probed by a pointed wire until a rectifying metal-to-semiconductor contact was obtained. The principal substances then being used were silicon and lead sulfide (in the form of galena). By 1909, Karl Baedeker was using the Hall effect to study systematically semiconductor behavior. J. Konigsberger published papers in 1907 and 1914, classifying silicon, selenium, and tellurium as semiconductors (germanium was not added until 1926).

The invention of the thermionic valve and its success as an amplifier soon overshadowed the crystal diode, finally displacing it in radio receivers by about 1926. Consequently, interest waned, although rectification characteristics were still being steadily improved by empirical means. However, a theoretical explanation for the phenomenon of rectification was still lacking. A major factor limiting progress in this respect was that the electrical properties of semiconductor materials are extremely sensitive to the introduction of minute amounts of impurities, and this was not realized at the time. Since varying amounts of impurities are present within unrefined semiconductor material, device characteristics varied considerably as a consequence. Even if the problem had been appreciated, techniques of crystal purification were not then sufficiently advanced to permit manufacture of material of a suitable quality in order to construct solid-state amplifying devices.

The introduction of the copper or copper oxide rectifier by L.O. Grondahl and P.H. Geiger (Germany) in 1927 was soon followed by the development of efficient selenium rectifiers, which were smaller and lighter for the same operating conditions, and could be operated at higher temperatures. Connected in series and provided with means of cooling, high reverse breakdown voltages were now possible. Their invention stimulated further interest in semiconductor rectifiers, since they extended existing commercial applications to include rectification in battery chargers and radios. It also raised interest in the possibility of controlling their current by means of a third electrode. Further uses for semiconducting devices followed the introduction of the selenium photovoltaic cell in 1931 by Bruno Lange and colleagues in Germany. This device operates by generating a voltage across a semiconductor junction when exposed to light, and was soon widely used in photographic exposure meters and applications such as control of artificial lighting, burglar alarms, and the opening and closing of lifts and doors. Photoconductive detectors, whose widespread use also dates from about this time, are devices that vary their conductivity when their surface is exposed to light. They have the advantage of low cost and quite high sensitivity, although their range of applications is restricted by their relative slowness in operation. A photosensitive material such as cadmium sulfide or cadmium selenide is usually deposited as a polycrystalline film onto a suitable substrate with a honeycomb of electrodes arranged to ensure maximum contact. These cells found an application as photographic exposure meters.

By 1930, little progress had been made in achieving a satisfactory theoretical explanation for the behavior of either semiconductor junctions or metal-to-semiconductor junctions, although a number of rival explanations were now being put forward, including an electrolytic barrier theory, the existence of a Peltier voltage generated by local Joulean heating, and also cold emission across a gap. Theoretical progress continued to be limited by the lack of reproducibility of experimental results, due to such factors as imperfections in crystal structure, bulk impurities, contact potentials, surface states, photoeffects and heating. Main techniques used in the investigation of semiconductor material at this time included Hall effect, conductivity, and thermoelectric power measurements. A further factor delaying development was that work in the field of solid-state devices was largely restricted to isolated individuals who lacked the resources necessary to mount a sufficiently large program of investigation.

Attempts were made in Germany, Britain, and elsewhere during the interwar years to construct various devices including field-effect transistors (FETs). However, these efforts met with little success. (In field-effect devices a current flows through a semiconductor channel, its width being controlled by means of a gate voltage). Probably the most well-known attempt was that made by Julius Edgar Lilienfield, then professor of physics at Leipzig, who took out patents in 1926 and again in 1928 for a voltage-controlled multilayer structure; one using a thin magnesium layer between

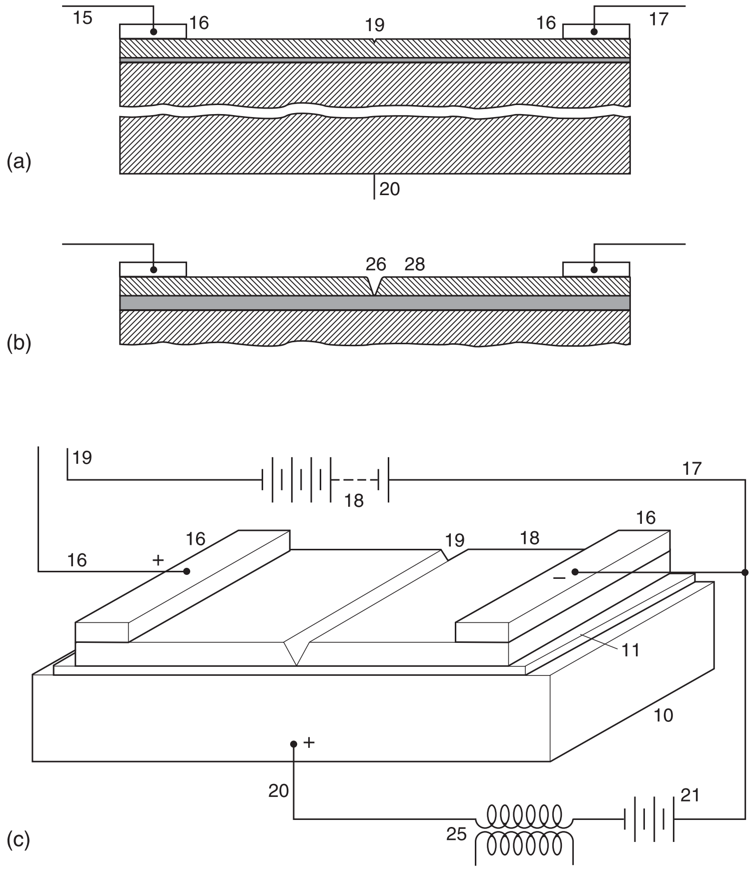


Figure 10. (a) and (b): The construction of a field effect transistor patented by J.E. Lilienfeld in the U.S. in 1933. The construction uses a thin layer of magnesium sandwiched between semiconducting copper sulfide layers. The notch shown in the center of (b) is intended to achieve a narrow cross-section, increasing the electric field at this point and hence the control effect. (c): This device shown with associated circuitry. [*Reproduced with permission from the Patent Office, U.K.*]

semiconducting copper sulfide layers (see Figure 10). There is no evidence that his devices actually worked. Nor did those of Oskar Heil (Germany) who in 1934 described a field-effect transistor with an insulated gate (see Figure 11). These patents were nevertheless important because their possession conferred priority. Rudolf Hilsch and Robert W. Pohl (Germany), working with alkali halide crystals in 1938, inserted a platinum control grid into the junction space-charge layer of a potassium

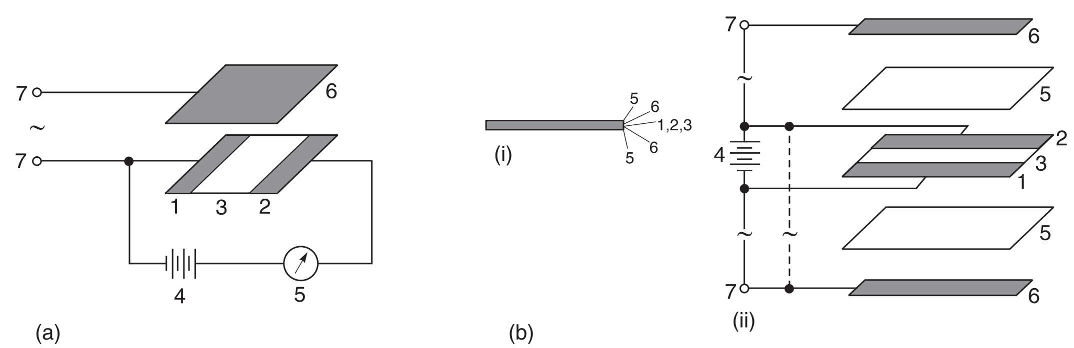


Figure 11. (A) and (B): The operation of a field effect transistor patented by O. Heil in Germany in 1934 and Britain in 1935. Heil states therein that ''one or more thin layers of semiconductor traversed by current is or are varied in accordance with control voltage applied to one or more control electrodes arranged to and insulated from said semiconductor layer or layers.'' (A) illustrates the principle of modulation, the current flowing within the circuit in series with the semiconductor layers being controlled by an alternating voltage applied to a control electrode situated above it. (B) illustrates the circuitry in more detail The unshaded plates shown in the illustration act as insulators.
[*Reproduced with permission from the Patent Office, UK.*]

bromide crystal (the width of which could be calculated), and modulating a signal by applying a grid voltage in similar fashion to that used in thermionic triodes. Alkali halides were chosen because large crystals of high purity could be obtained. However, the physical limitations of this approach prevented a useful frequency response being achieved, although their work demonstrated that it was possible to construct a solid-state amplifier. Further efforts to do so were made by William Shockley and Alan Holden (U.S.), working at Bell Laboratories in 1939, but little agreement was found between existing theory and experimental results. This was due to the lack of appreciation of the effect of minute amounts of unwanted impurities and other crystal imperfections, as well as the importance of the behavior of charge carriers at semiconductor surfaces.

See also **Photosensitive Detectors; Radio Receivers; Semiconductors, Compound; Semiconductors, Elemental; Semiconductors, Postband Theory; Transistors**

ROBIN MORRIS

Further Reading

Braun, E. and Macdonald, S. *Revolution in Miniature*. Cambridge University Press, Cambridge, 1982.

Braun, E. Selected topics from the history of semiconductor physics and its applications, in *Out of the Crystal Maze*, Hoddeson L.H., Ed. Oxford University Press, New York, 1992, 443–488.

Gibbons, M. and Johnson, C. Science, technology and the development of the transistor, in *Science in Context: Readings in the Sociology of Science*. Oxford University Press, Oxford, 177–185, 1982.

Gosling, W. The pre-history of the transistor. *Radio Electron. Eng.*, 43, 1/2, 10, 1973.

Loebner, E. Subhistories of the light emitting diode. *IEEE Trans. Electron. Devices*, ED23, 7, 675–697, 1976.

Morris, P.R. *A History of the World Semiconductor Industry*. Peter Peregrinus, London, 1990.

Sengupta, D.L., Sarkar, T.K *et al.* Centennial of the semiconductor diode detector. *IEEE Proc.*, 86, 23, 5–243, 1998.

Weiner, C. How the transistor emerged. *IEEE Spectrum*, 10, 24–35, 1973.

Wilson, A.H. Theory of electronic semi-conductors. *Proc. R. Soc. A*, 133, 458–491, 1931.

Wilson, A.H. Theory of electronic semi-conductors. *Proc. R. Soc. A*, 134, 277–287, 1931.

Ships, Bulk Carriers and Tankers

The era of the steam-propelled ship began in 1807 when Robert Fulton's *Clermont* paddle-wheeled up the Hudson River to Albany, New York. Steam application to ship technology progressed so rapidly that in only 30 years the *Great Western*, a 72-meter wooden paddle-wheeler owned by the Great Western Railway, was able to sail from Bristol to New York in 15 days and establish itself as the first steamship to perform regular service between the U.S. and Britain. The next seagoing milestone came in 1843 when the cargo ship *SS Great Britain*, with an iron hull, a state-of-the-art engine, and a propeller-driven thrust system, was launched in Bristol. Although it still carried masts and sails, the three elements of a modern cargo ship were present: iron hull, steam engine, and propeller drive. By 1914 the triple expansion reciprocating steam engine was developed and remained in use through the century. The first steam turbine for marine propulsion, built for demonstration purposes only, was Charles Parsons' *Turbinia* in 1894. Turbines were adopted for Navy ships such as *HMS Dreadnought* and in 1905 the first turbine-powered large ocean liner was built. The *Titanic* also carried a Parsons steam turbine engine. In 1912 the first cargo ship to be powered by an internal combustion engine was built. At the conclusion of World War I, some 300 "motor ships" were in operation. In addition to engine technology, improvements were made in hull design, cargo handling, fuel usage, and ease of repair.

The days of the wind-powered ship were over. Gradually the schooner and the sleek "clippers," represented by the *Cutty Sark* and *Flying Cloud*, were driven from the sea. The completion of the American transcontinental railroad ended the necessity of a 100-day ride around Cape Horn from New York to San Francisco, especially following the discovery of gold in California. The opening of the Suez Canal cut as much as 2500 kilometers off the sailing trip from China and India to Europe and the British Isles. In 1800, 2026 ships passed through the 162-kilometer ditch that connected the Mediterranean and Red Seas. The necessities of war in 1914 brought some of the schooners and smaller clippers out of "mothballs.," but they provided easy targets for submarine gun crews.

By 1914 the basic layout for dry-bulk carriers as well as oil tankers was established. Both were about 100 to 114 meters in length with a beam between 12 and 15 meters; metal hulled; equipped with a triple-expansion steam engine or possibly a steam turbine; were propeller driven and boasted no auxiliary masts and sails. The dry-bulk carrier had its housing set in the midst of five holds, three holds forward of the engine room and two holds to the rear. The two aft holds were penetrated by a

protective drive shaft housing which connected the engine via the drive shaft to an exterior propeller. About 11,000 cubic meters of cargo could be carried in these vessels. Within 30 years of Edwin Drake drilling the first oil well at Titusville, Pennsylvania in 1859, the first oil tanker was floated (1886). Dimensionally it differed little from the dry-cargo vessel. Safety dictated that the engine be set to the rear of the vessel so that no piercing of tanks and bulkheads was necessary for the propeller shaft housing. The ship's superstructure was set fore and aft with the hull tanks in the middle. Both bulk carriers could travel about 400 kilometers per day maximum; not much by twenty-first century standards.

The 1920s and 1930s were not decades of great prosperity. Some new ships were built and some old ships were refurbished, but no large construction orders were received until Henry J. Kaiser agreed to construct 2700 Liberty ships shortly after the outbreak of World War II. About 2500 Liberties survived the conflict. Eventually these were sold to a variety of purchasers, and only two were still afloat at the turn of the twenty-first century.

If a new ship is desired, the purchaser should know what he wants his ship to accomplish. A reputable naval architect should be contacted and his services engaged. The architect is responsible for determining the ship's design: size, shape, cargo capacity, and propulsion. The hull design likely would be determined by a 32-point dimensioning system utilizing computer-aided design (CAD) software. Ensuring the ship's buoyancy in water is essential. Displacement weight and volume has to be calculated so the vessel will float properly. Mathematical and computer models are used to work out the shape and displacement requirements and establish the ship's draft; that is, how deep the vessel sits in the water. Ship stability in the rolling sea, which can generate sideway and endway tilts, must be provided in the structural scheme. Resistant forces present when the ship is in motion must be countered by proper hull design and propulsion technologies. The ship will probably be furnished with a two-stroke diesel engine directly connected to the propeller shaft. The propeller will be either a fixed-pitch propeller or a newer controllable-pitch propeller. The fixed-pitch propeller has passed the test of time and is a solid design. It will be energized to the hull by a thrust block thus limiting its impact on the engine machinery.

The architect, with the help of a marine engineer, will combine the hull; superstructure; propulsion system machinery; ventilation, air, and heating apparatus; cargo handling equipment; and any other necessary gear into a profit-generating commercial vessel. A fine example of cost-ineffectiveness came after World War II with construction of the chic, sleek *NS Savannah*, a bulk and break-bulk carrier with passenger accommodation. It had 17,800 cubic meters of cargo space and could carry 9,400 tons of cargo at 20 knots. It required a large crew of more than 100 technically trained sailors, as it was powered not by a diesel engine, but by a nuclear reactor. It was launched in 1959 and retired 10 years later after sailing 730,000 kilometers and burning 74 kilograms of uranium ore. The cost of nuclear fuel and the salary requirements of a specially trained crew of 100 sailors made it very cost-ineffective when compared with a conventionally powered motor ship. It is now a deactivated museum piece in South Carolina. Nuclear powered vessels still have their adherents but there were no more Savannahs built.

One of the concerns of bulk carrier owners was speed and proficiency in the loading and unloading of cargo. New York shipbroker, Ole Skaarud, and a few collaborators dedicated themselves to improving the process. They increased the size of holds for better cargo handling and raised the hatches, especially for grain loading, several feet above the decks. Ship's housing and machinery were placed aft so that unobstructed holds could be installed in the central portion of the ship. Some vessels had cranes, booms, conveyors, and even pneumatic tubes attached to the deck while others relied upon off-ship loading facilities. A newly developed conveyor boom attached to a ship could swing over the holds and convey cargo off the ship. Skaarud also realized that overhead loading of bulk holds usually created a pointed load with the cargo sloping down to the hull sides and bulkheads. To improve safety and control cargo movements in rough seas, ballast tanks were installed in the upper reaches of each hold. Improved hull construction had lessened hatch failures, which in turn lessened breeched and flooded holds. Obviously, larger holds led to larger ships. There are four classifications for bulk carriers based on their dead weight tons (DWT) carrying capacity: Handysize, 10,000 to 35,000 DWT; Handyman, 35,000 to 50,000 DWT; Panamax, 50,000 to 80,000 DWT; and Capesize, 80,000 to 200,000 DWT. Panamax and Capesize cannot pass through the Panama Canal. Some bulk carriers move unusual cargo that requires special safety and ease of handling devices. These may be "reefers," with refrigerated cargos, multidecked and well-ventilated livestock carriers, and roll on/roll off (RO/

RO) ships which are designed to allow motor cars to drive on and off ships through doors in the hull. Many sailors consider the RO/RO carriers break-bulk vessels. The "banana boats" of fiction actually exist and service customers. The silhouette of the old cargo ship on the horizon has changed dramatically in the last half of the twentieth century.

No new development in sea traffic since the end of the World War II was more significant than the creation of containers and container ships. These were a product of intermodal transport invented by Malcolm McLean, a struggling truck, or lorry, operator in the Depression of the 1930s. He placed a motor truck trailer on a railway flat car destined for a seaport (or a railway terminal) where the trailer was removed to a ship (or was driven away). Soon the trailer was joined by a steel container 8 feet high by 8 feet wide (2.5 by 2.5 meters) and either 20 feet or 40 feet (6 or 12 meters) in length. The containers were placed on the ship, above and below deck, using standardized loading devices; a considerable difference from the old winch and crane loading system. Ships were soon designed and constructed specifically to carry containers. The first such ship, with 226 containers, set sail in 1957. Measurements in the ocean container trade are based on twenty-foot equivalency units (TEU). One TEU equals an 8 by 8 by 20 container. There are now many container ships that boast 6,000 TEUs; but it is prophesied that 12,500 TEU ships are not far in the future. Container ships are quite fast, ranging from 20 to 25 knots—much faster than tankers. Containers are offered for sale or lease in many varieties including dry bulk, break-bulk, liquid (tanktainer), insulated, and refrigerated. It is said that 10 percent of all sea traffic today is containerized, but this may be an understatement.

Large oil tankers have been described as floating behemoths. A few current tankers can carry 350,000 DWT of liquid cargo, usually crude or product (refined) oil. Gas tankers are essential where pipelines do not or cannot exist. Carrying natural, butane, or propane gas, these ships come in many sizes and are often easily recognizable by domed tanks rising above the ship's deckline.

Like container ships, tankers are classified by size. The largest are called ultra large crude carriers (ULCC) and run 350,000 DWT or higher. The inability of many oil ports to handle ships of this size and their high cost of operation has led, in recent years, to a substantial decline in their construction. Often they must be unloaded, via a floating pipeline, 2 or 3 kilometers offshore. The very large crude carriers (VLCC) are next in capacity with 200,000 to 350,000 DWT. These tankers are more port-acceptable, and thus more popular than the ULCC. They cannot, however, traverse the Suez Canal if fully loaded. It requires a Suezmax at 130,000 to 160,000 DWT to do this. New tankers, for safety, are now required by law to be double-hulled. Some tankers, classified FPSO (floating, production, storage, and offloading), anchor above an oil source and draw up the crude oil by pipeline hose from tanker to ocean floor. The 1967 closure of the Suez Canal did not cause great inconvenience because most ULCC and VLCC tankers are too large for this facility.

Europe and the Russian Federation offer good examples of efficient control of bulk traffic movement on rivers and canals. The use of barges and motor ships on inland waterways is more common in Europe than in the Americas. The Danube, Volgo, Rhine-complex, and Elbe Rivers are constantly dredged and kept in good repair. Interconnecting canals such as the Main Danube give the waterways system considerable reach and ease of movement. Barges and motor ships transport bulk cargos of coal, oil, ores, construction materials, fertilizers, and foodstuffs in, out, and through the system.

Barges are of various sizes, dependent upon the width and depth of the waterways in which they operate. Some are self-propelled, but most are pushed or pulled by towboats or tugs. Barges may be open, closed, tanked, or hoppered. Lighters can be seen—these are covered barges unloaded from an oceanic LASH (lighter aboard ship) vessel at a river, lake, or canal port and pushed or pulled by a tug or towboat to their final destination on the river, lake, or canal. After the cargo is unloaded, the procedure may be reversed.

The use of diesel or turbine-electric motor ships in Europe is also common, but take second place to barges in total tonnage moved. Motor ships range in capacity from 100 to 5,000 DWT, the latter with capabilities for movement in the Baltic, Mediterranean, and Black Seas. Most of these vessels resemble their oceanic colleagues, only smaller. They have as many as four holds, with machinery and housing in the rear and offer a speedier, but costlier, alternative to the slower barges.

The Great Lakes in the U.S. offer a variety of clumsy-looking ships, up to 300 meters in length, with many holds, and topped off with superstructures fore and aft. Some possess very modern self-unloading equipment. They mingle with pleasure craft, sailing vessels, ocean-going tankers,

bulk, and break-bulk ships. The latter, called "salties," enter the Lakes through the St. Lawrence Seaway. The Great Lakes freighters, or "longships" are 150 to 300 meters in length and move around the Lakes at 10 to 15 knots carrying as much as 60,000 tons of iron ore from the Mesabi Range in Minnesota, coal from Appalachia, or grain from the Great Plains. Any cargo ship that moves in and about the Lakes must fit into the locks at Sault St. Marie, Michigan or the locks in the Welland Canal system in Ontario. "Salties" cannot ascend Niagara Falls without utilizing the Welland Canal. As a result of size some "salties" get only this far. The "longships" and "salties" sometimes travel in very rough weather.

Lloyd's Register of Shipping lists over 85,000 vessels of more than 100 gross tons. No small group of nautical corporations can dominate the shipping trade, for they lack sufficient vessels for control. The Nippon Yusen Kaisha (NYK Line) may be the world's largest seagoing cargo carriers. It has three container groups, two bulker groups, and three petroleum and gas groups. Nevertheless, NYK can float no more than 560 ships. The Maersk SeaLand and SafMarine companies in the A.P. Moller Group (Denmark) total 290 container vessels, the world's largest inventory of these ships. It is interesting to note that Moller acquired this unit—Maersk SeaLand—in 1999 from CSX Corporation, an American railway conglomerate. P&O Ned Lloyd Container Line Ltd. (U.K.) boasts the second largest container fleet with about 146 vessels. Many large seagoing carriers are, in fact, conglomerates with divisions and subsidiaries outside the shipping industry.

The greatest growth in the shipping industry is in container vessels. In the past two years dry-bulk and oil carriers have remained quite stable; that is, no growth and no loss. Container ships appear to have the most promising future.

ERNEST M. TEAGARDEN

Further Reading

Armstrong, R. *The Merchantmen*. Praeger, New York, 1969. Fine work on steamship development.

Batio, C. *Super Cargo Ships*. MBI Publishing, Osceola, WI, 2001.

Hornsby, D. *Ocean Ships*. Ian Allen, Shepperton, Surrey, 2000.

Know Your Ships. Marine Publishing, Sault Ste, Marie, MI, 2002.

Labaree, B.W. *et al. America and the Sea: A Maritime History*. Mystic Seaport Museum, Mystic, CT. 1998. A 620-page history of seafaring Americans from colonial times to the late 1990s.

Nersesion, R.L. *Ships and Shipping: A Comprehensive Guide*. Penn Well, Tulsa, OK, 1981.

Thompson, M. *Queen of the Lakes*. Wayne State University Press, Detroit, 1994.

Valkar, I. Investment in inland waterways? Infrastructure needs. *Seminar on the Inland Waterways of Tomorrow on the European Continent*, Paris, 30 January, 2002. Document No. 5.

Useful Websites

Nuclear Merchant Ships: http://www.atomicengines.com/ships.html.

Ship Design and Construction: http://www.mi.mon.ca/mi-net/shiptech/shipbldg.htm.

Ship's Propulsion Layout: http://members.shaw.ca/diesel-duck/machinery_page/propulsion_layout/propulsion_layout.htm

Skyscrapers

Skyscrapers are the world's tallest buildings. One-hundred-eighty- and 200-meter-high buildings that were considered to be exceptionally tall in 1910 were overshadowed by skyscrapers of more than 300 meters in a matter of 20 years. Advances in construction techniques enabled engineers to build ever-taller structures throughout the twentieth century. However, the principle reasons for erecting exceptionally tall buildings changed little over time. Densely populated cities with escalating land values called for maximum utilization of available space, and tall buildings are one of the most economical means of assembling large numbers of workers in one place. While the majority of skyscrapers were built for the profits they could generate, other reasons included self-aggrandizement, prestige, image, and recognition.

The skyscraper is a late nineteenth century American innovation that was more aptly defined by its height than any characteristic of construction or particular style. While the genre first appeared in Chicago in the 1880s, New York City became the focal point of tall building construction shortly after 1900 and remained so for much of the twentieth century.

The increase in the height of New York's tall buildings was rapid. One of the first exceptionally tall structures was the 187-m, 47-story Singer Building erected in 1906. In 1912, the 241-meter-tall Woolworth Building boasted 60 stories. The height of the tallest had vaulted to over 300 meters in 1930 when the Chrysler Building was completed. Only a year later its height was eclipsed by the Empire State Building, standing 381 meters high.

At that point it was as if an upper limit had been reached and the Empire State Building remained the world's tallest building for over 40 years. There were no extraordinary increases in height in the world record buildings that followed and they tended to be only fractionally taller than the previous record holder. In 1972, the 417-meter twin towers of the World Trade Center were completed but their supremacy was brief as Chicago's Sears Tower, standing 19 meters taller, opened in 1974.

Despite the burgeoning technology of tall building construction in America, Europe did not follow suit. Five- or six-story buildings seemed to be the limit there until after World War II. It was not until 1990 that the first large-scale skyscraper, the 257-meter-tall high-strength concrete Messeturm, was constructed in Frankfurt, Germany. Although the situation differed little in Britain, efforts were made in London to impose height limits on new buildings out of concern that historic buildings not be overshadowed. However, at the end of the twentieth century there was active planning for a number of skyscraper-size buildings.

Structurally, most skyscrapers were basic steel frame buildings assembled of columns and beams. Columns passing through the interior of the structure from bottom to top carried the structural load down to the building's foundation. Beams supported both concrete floors and lightweight curtain walls of terracotta, brick, glass, or metal. The first major exception to this configuration was the hollow tube design of the towers of the World Trade Center in which the outer walls were load bearing. The design featured a perimeter of closely spaced columns and the elimination of all but a main core of heavy internal columns created large open floors. This same form of tube of construction was used to build the Sears Tower, which until 1998 was the world's tallest building.

Throughout much of the century, hot rivets were used to assemble the structural steel members of skyscrapers. During the 1940s, the first of several significant technical developments occurred beginning with the increased use of welded joints. Yet another important advance came about during the 1950s when high-strength bolts were introduced for use in connections. The ductility of joints consisting of a combination of both welding and bolting made them so superior to existing methods that by the early 1960s riveting had been all but abandoned for building construction. Welded and bolted frames could be further enhanced with eccentric bracing that better resisted the lateral forces of wind pressure and earthquakes. Earthquake-resistant building design continued to be one of the most complex and challenging problems faced by engineers. Possible solutions included the use of shear walls designed to resist sideways forces and the isolation of a structure from its foundation by separating the two with layers of rubber.

From the 1950s, new high-strength concrete provided engineers with a reasonable economic alternative to traditional steel frame construction for tall buildings. Without compromising strength, columns could be smaller and in turn they freed up rentable floor space. Nonetheless, concrete structures had higher mass and a greater damping effect in wind loads and movement caused by earthquakes, and they outperformed their structural steel counterparts. Concrete had the added advantage of being both fire and blast resistant.

Skyscrapers require the same systems for heating, air conditioning, lighting, and power that are used in their shorter counterparts; however, while the installation of most systems pose no particular problems, supplying water to tall buildings does require special equipment. As typical city mains pressure would raise water only a few stories, pumps are needed to move water to integral storage tanks which in turn supply the structure. However, of all the systems that go into a skyscraper, none is more critical than the elevator. With five or six stories being the maximum that can be reached readily by foot, the elevator is essential to the utilization of a major part of a skyscraper. Bulky power sources of early elevators typically took up valuable floor space. The system was revolutionized in 1903 with the introduction of electric gearless traction and an elevator's motive power could be located at the top of the shaft.

In the early 1990s, the center of tall-building construction moved from North America to Asia where there was great activity in building skyscrapers. The century of the skyscraper was brought to a close in 1998 with the completion of the world's tallest buildings, the two 452-meter-high concrete Petronas Towers in Kuala Lumpur, Malaysia.

The destruction of both towers of New York City's World Trade Center by terrorist attacks in 2001 raised anew questions regarding the desirability and safety of skyscrapers as a building form. Technologically, the upper limit to the height to which buildings may rise has yet to be reached. Although economics remains key among the factors deciding how high buildings rise, it is

politics that may play a pivotal role in determining if towering skyscrapers are built in the future.

See also **Fire Engineering; Vertical Transportation (Elevators)**

William E. Worthington, Jr.

Further Reading

Tauranac, J. *The Empire State Building: The Making of a Landmark*. Scribner, New York, 1995.

Landau, S.B. and Condit, C.W. *Rise of the New York Skyscraper, 1865–1913*. Yale University Press, New Haven, 1996.

Goldberger, P. *The Skyscraper*. Alfred A. Knopf, New York, 1981.

Beedle, L.S., Ed. *Second Century of the Skyscraper, Council on Tall Buildings and Habitat*. Van Nostrand Reinhold, New York, 1988.

Douglas, G.H. *Skyscraper: A Social History of the Very Tall Building in America*. McFarland, Jefferson, 1996.

Elliott, C. *Technics and Architecture: The Development of Materials and Systems for Buildings*. MIT Press, Cambridge, MA, 1992.

Smart and Biomimetic Materials

Smart materials react to stimuli such as heat, light, electricity, or some other factor from their external environment. They include shape-memory alloys and polymers, piezoelectric polymers and ceramics, biomimetic polymers and gels, conductive polymers and controllable fluids, magnetostrictive materials, and chromogenic materials. Smart materials can be used in a smart materials system, such as micromachined electromechanical systems and fiber-optic sensor systems. Smart materials systems are a late twentieth century design methodology that integrates the actions of sensor, actuator, and control circuit elements in systems that can respond and adapt to changes in their environment or condition in a helpful way. They bestow smartness, or a functionality that enhances the value of materials, technologies, or the final products. Smartness can improve the performance of a system in a way that cannot be achieved using the more established non-smart approaches.

Many other smart materials systems are continually researched and developed, notably by the defense and aerospace industries. The National Aeronautics and Space Administration (NASA), for example, is developing the ultimate smart structure: a shape-changing airplane wing that will reorder and optimize its shape during flight to match the atmospheric conditions or the mission it must carry out.

Shape-Memory Materials

Shape-memory materials behave in a predetermined way when exposed to a particular stimulus. They can exist in different shapes at two or more different temperatures once their transition temperature has been reached. Shape-memory alloys have very different shape-changing characteristics compared with shape-memory polymers. Devices made from shape-memory alloys can provide force when they are exposed to their transition temperature, as is the case with actuators. However shape-memory polymer devices undergo mechanical property loss while exposed to their transition temperature, as when used to make releasable fasteners.

Shape-memory alloys are much more costly than polymers. It takes time to program a metal alloy that also needs heat treatment at temperatures of hundreds of degrees Celsius, resulting in a maximum deformation of only about 8 percent. Shape-memory polymer systems are being developed that are easier to shape, with more applications than the shape-memory materials already in use, such as the nickel–titanium alloy Nitinol, used to make items such as flexible spectacle frames. The polymers can be programmed into shape in seconds at about 70°C and can undergo major deformations of several hundred percent.

Fiber-Optic Sensors

Fiber-optic sensors are based on optical fibers attached to sensing devices, forming smart systems that may be built into structures such as airplane wings and bridges. Fiber-optic sensor technology is used to measure mechanical properties such as strain, pressure, and temperature. Structures incorporating fiber-optic sensors react to or warn of imminent failure and demonstrate the state of the structure following damage.

Piezoelectric Polymers and Ceramics

Piezoelectric polymers and ceramics are materials that exhibit new properties when they are exposed to an electric current (piezo-electricity). Many polymers, ceramics, and molecules such as water are continuously polarized, with some parts of the molecule being positively charged and other parts negatively charged. By applying an electric field to these materials, the polarized molecules will align themselves with the electric field, producing induced dipoles within the molecular or crystal structure of the material.

Piezoelectric materials are used in acoustic transducers that change acoustic (sound) waves

into electric fields and electric fields into acoustic waves. These find an application in devices such as speakers and drums.

Semicrystalline polyvinylidene fluoride (PVDF) has been the only piezoeletric polymer commercially available until recently. Lightweight, flexible and easily produced as sheets or in complicated shapes, its low mechanical and acoustic impedance makes it highly suitable for use in underwater and medical applications. However PVDF has limited temperature use and poor chemical stability in extreme environments. Polyimides may be an alternative as they have excellent thermal, mechanical and dielectric properties combined with high chemical resistance and stability.

Magnetostrictive Materials

Magnetostrictive materials can convert magnetic energy into mechanical energy and also transform mechanical energy into magnetic energy. Magnetostriction is a property of the material that does not lessen over time. These materials expand when exposed to a magnetic field, an effect known as the Joule effect or magnetostriction after James Prescott Joule. In the early 1840s Joule identified the phenomenon when he observed a change in length of an iron sample as its magnetization altered. This effect is due to the lining up of the magnetic domains in the material with the magnetic field. A change in the size of the width occurs together with the change in length produced by the Joule effect. When the material is stretched or compressed, it undergoes strain and its magnetic energy changes. This is known as a magnetomechanical or Villari effect, commonly used in magnetostrictive sensors. Magnetorestrictive materials include iron, cobalt, nickel, ferrite, metglass, and terbium alloys (Terfenol-D). They are used in a range of modern devices such as sensors, sonar and ultrasonics, speakers, vibration and noise control, and drills and reaction mass actuators.

Chromogenic Materials and Systems

Chromogenic materials and systems can change their optical properties in response to an electrical, photo, or thermal stimuli. Electrically activated chromogenic systems are used for smart windows and mirrors in the automotive and architectural industry and for low-information content displays. Electrically activated chromogenics can be controlled by the user, unlike photochromic and thermochromic devices which are self-regulating and passive.

Electrochromic materials are chromogenics that change color on electrical stimulation. They are actuator elements that need sensor and control circuits to be added to make the system smart. Reflective hydrides may also be regarded as electrochromics, but they differ in a number of ways from the more commonplace oxide electrochromics. Originally deposited as a metal, they can be converted to a partially transparent hydride by injection of hydrogen from the gas or solid phase when they switch to a reflective state, which has several potential advantages in terms of energy performance and durability. Transition-metal hydrides have now also been developed. Liquid crystal windows switch quickly from a transparent state to a diffuse white state; however, they have little control over solar heat gain. Suspended particle displays are also under development. Photochromics darken in sunlight and are therefore mainly used to make sunglasses that darken automatically. Thermotropic materials respond primarily to heat.

Heat-sensitive polymers (thermochromic) are used for children's toys, tee-shirts, and toothbrushes that change color when touched. These materials have additional functions that are visible in real time and can provide a variety of intelligent responses.

Biomimetic Materials

Biomimetic materials are based on nature's best designs and attempt to mimic them. Nature solves problems by looking for a solution that works using the minimum amount of energy. Engineers solve problems by searching for an effective solution with the lowest cost. Plants and animals require a great deal of energy to produce the basic materials they need for survival, but they can use almost any shape. Engineers are able to produce a wide variety of materials cheaply, but shapes are often expensive to make. By copying nature's designs and shapes, researchers can make more efficient structures that can be used to solve tomorrow's engineering problems as well as to develop innovative new materials. Biomimetic polymers are being developed based on natural materials; for example, spider's silk, which is in fact a biopolymer. Biomimetic gels; for example, those based on the sea cucumber, are being researched and may result in a number of biomedical applications.

Microelectromechanical Systems (MEMS)

Microelectromechanical systems (MEMS) are miniaturized devices. They may be as small as a silicon

semiconductor chip, which is able to integrate sensors, information or signal processing, and control circuits in a single device. A range of MEMS -based devices have been developed including pressure sensors, transducers, transmitters, microrelays, optical attenuators and photonic switch components, and smart security and tagging systems. They are under development for further biomedical applications.

Controllable Fluids

Controllable fluids have properties that depend on an electric or magnetic field. They are a smart technology that may be used instead of piezoelectric transducer-controlled semiactive suspension systems.

Conclusion

Among the more interesting biomedical developments in smart systems in the 1990s were synthetic muscle actuators, which included shape memory alloys, piezoelectrics, and electroactive polymers. A particular example of a smart material used for such an application is IPMC (ion-exchange polymer membrane metallic composites). The synthetic muscles contract or bend when exposed to an electric current and can be made into wires that are as thin as a human hair.

See also **Ceramic Materials; Photosensitive Detectors; Plastics, Thermoplastics**

SUSAN MOSSMAN

Further Reading

Friend, C. *From Sensuous Aeroplane to Cuddly Toaster*. Institute of Materials, London, 1997.

Culshaw, B. *Smart Structures and Materials*, Artech House, London, 1996.

Holnicki-Szulc and Rodellar, J., Eds. *Smart Structures: Requirements and Potential Applications in Mechanical and Civil Engineering*. Kluwer Academic, London, 1999.

Srinivasen, A.V. and McFarland, D.M. *Smart Structures: Analysis and Design*. Cambridge University Press, Cambridge, 2000.

Vincent, J.F. Stealing ideas from nature. *RSA Journal*, August/September, 36–43, 1997.

Vincent, J.F. Naturally new materials. *Mater. Today*, 1, 3, 3–6, 1998.

Social and Political Determinants of Technological Change

The question of whether and to what extent social and political factors determine or influence technological change is a major issue in the historical and social understanding of technology, one that took special form during the twentieth century. Precisely because technological change has been so persistent and so closely associated with obvious social and political transformations, questions have regularly been asked about both the drivers of the inventive creativity and the parameters of its social and political interactions.

Focusing on the U.S., which has often been a site for the leading edge of such interactions over the course of the twentieth century, one may distinguish three major periods of technological change.

First, the century opened in the presence of emerging industrial conglomerates controlled by powerful families such as Carnegie, Rockefeller, and Vanderbilt and the rise of corporations increasingly committed to technological research and development such as Bell Telephone, Eastman Kodak, Edison Electric, Ford, and General Motors. These new social constellations appeared in conjunction with new technological systems of unprecedented scale and complexity. Coal and iron ore mines were linked by railroads with smelters and mills to make steel for use in the assembly-line production of automobiles and other goods to be consumed by an increasing number of people in urban centers, the populations of which for the first time surpassed those of rural areas.

During the broad middle third of the century, post-Depression and especially in conjunction with World War II, technology became increasingly science dependent and rationalized. Examples ranged from physics for the development of radar, the atomic bomb, and computers; to chemistry for the creation of dyes, synthetic, and agricultural chemicals; biology for hybrid seeds and antibiotics; and aerodynamics for the design and manufacture of airplanes and rockets. At the same time as the government established a system of national laboratories, research and development activities were increasingly promoted on university campuses. The need for more efficient industrialization and effective management of large human–machine complexes such as battle groups incorporating naval, army, and air force components stimulated the development of operations research and analytic management techniques well beyond traditional accounting and marketing procedures. After the war, the reconstruction of Europe and Japan, the rise of international capitalism (Bretton Woods, the World Bank, and the International Monetary Fund) and politics (the United Nations, the Cold War, and Third World development programs) was further associated with extensions

of technological change as consumer goods multiplied in availabilities and technical sophistication.

A final short third of the century witnessed the emergence of sometimes competing criticisms of various aspects of these massive technological and social developments. The environmental movement and technological disasters (Three Mile Island, Challenger, Chernobyl, Bhopal) raised questions with respect to any simple assumptions about the unqualified beneficence of technological change. The antinuclear, consumer, anti-Vietnam War, and AIDS activist movements called for greater public participation in technical decision making and reorientations in social priorities. Economic and political assessments of bureaucracies challenged the efficacy of large-scale hierarchical organizations and regulatory agencies. Together with the energy shortages during the 1970s and the end of the Cold War in 1989, environmental and economic analyses were associated with transformations toward flexible systems of production distributed across many countries, designer consumer goods with niche marketing, and the promotion of new electronic communications, from computers and cable or satellite television to the Internet and cell phones. The century closed with the rise of interactive digital technologies that created virtual realities, and of biotechnologies and nanotechnologies that challenged traditional boundaries between the natural and the artificial.

Theories of Science, Technology, and Society Interaction

Across the twentieth century technological change appeared, paradoxically, to be both inevitable and the result of the free choice of human beings. The common belief has certainly remained that sociopolitical factors do in fact influence both the direction and rate of technological change. Precisely for this reason, early twentieth century historians sought to explain the singular rise of modern technology in Europe since the 1500s by appealing to the influence of social and political factors such as agricultural productivity, resource availability, population, economic structures, religion, or democracy; while sociologists and economists sought to specify the circumstances under which technology could effectively be transferred from one context to another, either from one country to another or from laboratory to market. At the same time, people in the twentieth century sometimes experienced technology as functioning in a semiautonomous manner, determining or influencing more than being influenced by society, politics, or economics. In response, governments have often looked for ways to control and regulate technological change, but have not always found themselves successful in doing so.

Ironically, the notion of technology as a semiautonomous force, itself determining social and political phenomena, has been at once hailed as good and criticized as bad. The motto of the Chicago Centennial of 1933 celebrating "A Century of Progress" was "Science Finds, Industry Applies, Man Conforms." During this same period, however, University of Chicago sociologist William Fielding Ogburn coined the term "cultural lag" to describe disharmonies that resulted when technological change outstrips social and political development. One of his examples was how technological increases in productivity and medically promoted decreases in infant mortality combined to undermine the rationale for large families, although it normally takes one to two generations for family size to drop appropriately. Attempts to clarify the precise senses in which technological change is and is not an independent variable in relation to society and politics was thus a recurring theme in twentieth century reflections on technology.

The bottom line is that there exists no comprehensive theory of the ways technology determines or influences society and politics or of the ways society and politics influence technology. Indeed, there is not even any consensus about how to define technological change, which because of its complexity is often termed "technosocial" or "sociotechnical change"—thus finessing rather than answering the theoretical question. This makes it impossible to review the phenomenon of technology–society relations themselves; the best that can be done is to survey theoretical approaches to such relationships. From this perspective, scholars during the first half of the century tended to concentrate on searches for macro theories about technological influence, while those of the second half stressed micro theories about how social and political factors determine or influence technology. Near the close of the century scholars began advancing theories that attempted to bridge the two positions, but without fully satisfying the challenging complexities of the multiple questions at issue.

Theories of Technological Determinism

Theories emphasizing the influence of technology on society and politics, often termed theories of technological determinism, inherited from the

nineteenth century three basic stories about the history of technology. These three competing and collaborating stories are

- The idea of the progressive human conquest of nature (insofar as technologies are well-defined extensions of human action).
- The massive industrial proliferation of goods and services (through increased division of labor and standardized production routines).
- A tendency of technology to itself escape human control (as artifacts progressively tap non-human energy resources, formalize human behavior routines, and overwhelm the social world with the products of human labor that produce manifolds of unintended consequences).

The work of American cultural critic Lewis Mumford and French sociologist Jacques Ellul explored, refined, and criticized these three inherited stories, and in the process presented theories of technology as an underappreciated and often unwelcome determinant of many aspects of society and politics.

Mumford's *Technics and Civilization* (1934) took the anthropologist's periodization of human prehistory as defined by materials (Stone Age, Bronze Age, Iron Age) and extended it into history, proposing to complement the more traditional cultural and political distinctions (Greek and Roman Ages, Middle Ages, modern period of nation states) with one defined by changes in technologies. Mumford's proposal was to distinguish the "eotechnic" use of animal power and increasingly specialized tools during the Middle Ages, the "paleotechnics" of steam power and the machines of the Industrial Revolution, and the "neotechnics" of electric power in the twentieth century, each sponsoring special social and political orders. Although his terminology never took hold, the book itself virtually created the notion of a social history of technology, all previous attempts being more internalist technological histories. In one of his vivid illustrations of the power of technologies to influence social order, Mumford argued that the invention of the clock transformed the experience of time; before the clock, for instance, working days had varied in length with the seasons of the year. The clock made possible the standardized working day and thus the mass production factory in which all workers for a particular shift are required to show up at the same time. A later example was nuclear energy, which virtually required centralized, hierarchal, anti-democratic management.

For Mumford the twentieth century was in a situation comparable to that of the Industrial Revolution, having created powers that it now needed to learn to guide and manage. Making alliances with socialists and liberal progressives, he called for a more conscious appreciation of this circumstance and an effort to bring a broad spectrum of human values to bear on what he saw as a monomaniacal technological world focused too exclusively on the pursuit of power. In his last work, however, *The Myth of the Machine* (2 vols, 1967 and 1970), Mumford appeared to despair of the possibilities for this exercise in human responsibility, although in reality his arguments almost certainly had an influence on the environmental and economic criticisms of technology.

Whereas Mumford's intellectual background was literary-based cultural criticism, Ellul sought to extend perspectives inherited from Karl Marx. Even though he was not a strong technological determinist, Marx credited technology with exercising a powerful social and political influence. In a famous phrase from *The Poverty of Philosophy* (1847), Marx remarked, "The hand-mill gives you society with the feudal lord, the steam-mill society with the industrial capitalist." Of course, one then has to ask, but what gives you the technologies of hand-mill or steam engine. To claim, as Marx did, that such technological change is a more or less unconscious process of human interaction with the natural world and existing technologies, which then have unanticipated social consequences, constitutes a kind of soft determinism.

This is the view developed at length by Ellul in *La Technique* (1954), which begins by distinguishing between "technical operations" and the "technical phenomenon." Polytechnical operations are present throughout history, always subordinate to the contexts in which they occur. The technical phenomenon, by contrast, is a unified method for the guidance of making and using that arises more or less by accident. However, once it comes on the historical scene, this phenomenon begins to transcend all particular contexts with appeals to decontextualized notions of efficiency. The public faith that human beings then place in "technological efficiency" furthers its dominance. To appreciate the determining power of this new phenomenon, which may also be described as technology turned into a social institution, requires a special act of consciousness. It is Ellul's aim to contribute to the emergence of this consciousness with what he calls a "characterology" of technology.

According to Ellul's characterology, technology in the mid-twentieth century exhibits the distin-

guishing features of artificiality, self-augmentation, universality, and autonomy. It replaces the natural milieu with one increasingly fabricated by human beings, thus occluding natural orders and influences. It repeatedly extends itself ("The only solution to the problems of technology is more technology"). It is more and more the same everywhere ("If you've seen one Holiday Inn, you've seen them all"). It becomes increasingly independent of external values ("Why do something in a way that is inefficient?"). By means of these characteristic features, technology transforms economics, politics, medicine, education, sports, and entertainment, driving all such traditional human activities to incorporate technical means and seek efficiency in their respective areas of activity.

These two versions of technological determinism emphasized, in turn, the primacy of technology as physical artifact (Mumford) and human process (Ellul). In a more abstract interpretation, the German philosopher Martin Heidegger (1889–1976) argued that modern technology was constituted by a stance at once willful and epistemological; technology takes up an attitude toward the world that forces it to reveal itself as resources available to be controlled and manipulated. From each perspective—artifact, activity, will, and knowledge—technology is argued to exercise deep influences on politics and society. Langdon Winner's *Autonomous Technology* (1977) and *The Whale and the Reactor* (1986), along with a Merrit Roe Smith and Leo Marx edited volume titled *Does Technology Drive History?* (1994), provide fitting reviews of the determinist tradition in the midst of the emergence of a major alternative research program called social constructivism.

Theories of the Social Construction of Technology

During the latter half of the twentieth century, a new generation of scholars turned away from both grand theories and the idea of technological determinism in order to explore with case study detail the myriad ways in which social and political factors can, and indeed do, influence technology. In part, this change resulted from the conceptual maturity of academic programs in science and technology studies (STS), programs that first arose more or less simultaneously and independently during the 1970s in the U.S., Europe, and other parts of the world. Within this context, two leading proponents of the new contextualized approach that promoted theories of the social construction of technology were the Dutch historian Wiebe Bijker and the French social scientist Bruno Latour.

In an influential study of technological change in the bicycle, Bijker (1987) argues that this artifact never played the dominant role in all the social changes with which it was associated that technological determinism might have implied. The early bicycle, with its large front wheel and small back wheel, was quite unstable and known as a "bone breaker." Young men rode it in the park to display their masculine daring. The development of the "safety bicycle" occurred in conjunction with women asserting their right to appear in public places doing many of the things that men do, but maybe not just like men. The fact that boys still refuse to ride a drop frame "girl's bike," despite its structural superiority, clearly shows how social interpretation dominates technological change.

For Bijker the social meaning of an artifact is always underdetermined by the artifact itself. Technologies are thus subject to interpretative flexibility, and function as sites of competing adaptations. The history of technologies is a history not just of hardware, but of social contests about the meanings attached to the hardware and how it is going to be deployed in society. In Bijker's hands, social constructivism shows how different social groups (e.g., designers, regulators, users) negotiate the meaning and function of a technology. According to Bijker:

> "The sociocultural and political situation of a social group shapes its norms and values, which in turn influence the meaning given to an artifact" [*Bijker 1987, p. 46*].

However, Bijker also recognizes that one interpretative stratagem has been to describe technology as determining of social and political institutions. In an effort to account for this experience of consumer alienation in which technology is viewed as independent and out of human control when analysis reveals extensive dimensions of human construction, Bijker (in 1997) simply proposed that people with low inclusion in the construction process are often faced a "take-it-or-leave-it" choice when new technologies come on the scene. Technologies appear to them predetermined and autonomous because they had no part in the determination; but those with high inclusion in the design and construction process—engineers and trend-setting consumers—are quite conscious of their constitutive powers and thus do not feel so alienated.

In a complementary theoretical stance derived from the social sciences, Latour set forth an actor-network theory of technological change. Since the

middle part of the twentieth century, social scientists had been complementing functionalist and structuralist analyses of society with descriptions of actor networks. The character of a social institution such as a church or school was fully explained neither by its social function nor its structural features. The network of people involved in the institution was often of complementary importance. In a similar manner, Latour argued in *Science in Action: How to Follow Scientists and Engineers through Society* (1987) that what is central to understanding technologies is to see the relations between the actors involved. For actor-network theory, the settlement of controversies between actors over the final design of a technological artifact is the result of an effective deployment of allies and resources behind the winning design.

The often appealed to search for efficiency never resolves any design issue by itself, because efficiency can mean multiple things in multiple contexts. A car that uses less gasoline but fails to attract consumers is not an efficient product for the corporation that manufactures it. Through the positioning and deployment of other actor and networks, engineers constantly (re)create society ("society in the making") as they struggle to legitimate their designs. In *The Golem at Large: What You Should Know about Technology* (1998), two colleagues of Latour apply this perspective to a series of case studies—from debates about Patriot missile defense and the Challenger Space Shuttle disaster, and questions concerning nuclear fuel flasks and the origins of oil to economics, Chernobyl, and AIDS treatments—noting in each situation how engineers worked in diverse ways to enroll various interests behind different interpretations of technologies. For social constructivists the message is always that things could have been otherwise, and the need to recognize differences in situations where determinists and others tend to see only similarities and uniformity.

One way to summarize the upshot of social constructivism is to describe it as an attempt to deconstruct the inherited stories of technology as a progressive conquest of nature, increased production of goods and services, or a semiautonomous force, in order to write a new story with multiple authors seeking coherent narratives. Sometimes a narrative strategy succeeds in enrolling others, sometimes not; but whether successful or not, the contributors keep writing. What exists in the history of technology for this complex interactionist model is not so much progress or regress as sideways slipping and sliding; yet one cannot help but wonder why the notion of determinism persists to be argued against, or why the authors of technologies continue to write.

Beyond Determinism and Constructivism

Three other theories about relationships between technology, society, and politics that call for mention are those involved with feminism, evolution, and economics. Feminist theories, as introduced in Judy Wacjman's *Feminism Confronts Technology* (1991), can be related to both determinist and social constructivist perspectives. Insofar as women are excluded from the technological design process, some feminists argue this creates for women a historically contingent determinism. But what is desirable are situations in which women and their perspectives are allowed to contribute to the social construction of technology.

The idea that technological change can be understood as an evolutionary process has exercised persistent appeal. The most extensive development of this perspective is found in John Ziman's edited volume, *Technological Innovation as an Evolutionary Process* (2000). Ziman and colleagues explore the possibilities for analyzing technological change in terms of the selective retention of invented variations in products and processes, although they note that with technology neither variation nor selection is as blind as in the organic world.

The Ziman research program is dependent on a distinction initially developed by economist Joseph Schumpeter between invention and innovation, technical creation, and its economic exploitation. Yet even more important in Schumpter than this trope is his critique of equilibrium economics in the name of disequilibrium economics, the latter of which is characterized by an internal commitment to technological change. An economy so intimately involved with technological change becomes characterized by repeated phases of "creative destruction": water power, textiles, iron (early 1800s), steam, railroads, steel (late 1800s), electricity, chemicals, internal combustion engine (early 1900s), petrochemicals, electronics, aviation (mid- to late 1900s), digital networks, software, new media (late 1900s and early 2000s). As Marx and Engels noted in *The Communist Manifesto* (1948), in such a world "everything that is solid melts into thin air." For later economists, however, the thin air had become itself a phenomenon to be analyzed by rational choice theory, as Jon Elster's influential *Explaining Technical Change* (1983).

Conclusion

The twentieth century witnessed major changes in technology as well as in social and political affairs. Examples range from the creation of technological power in both economic and military forms associated with the U.S. becoming the dominant world power, through the technocratic industrialization of countries such as Spain under Franco that may have set the stage for later democratic developments, to the information technology revolution in the Soviet Union that (according to interpreters such as President Ronald Reagan) contributed to its complete collapse between—not to mention the industrial and information technology transformations in Japan, China, and India that can be correlated with wide ranging sociopolitical transformations. By contrast, the failures of technological development in much of Africa and significant parts of South America appear to be aligned with differing degrees of social and political failure.

Looking back at attempts to comprehend the creative destructions of such scientific, technological, and social interactions, scholars during the first half of century appear more concerned with how societies became technological in the first place, how technology burst into history and was then passed from one society to another. Scholars during the last half of the century, living in the triumphal wake of a new order of artifice, aspire to a more internalist comprehension of existing processes. Their case studies focus on situations that are themselves already highly technological and undergoing continuous technological transformations; and whereas macro theorists were often animated by grand issues of social and political justice, micro constructivists localized such questions and then turned them over to applied ethicists. The hope was that by acting locally one might have global impact, in ethics and politics as much as in technology.

CARL MITCHAM AND JUAN LUCENA

Further Reading

Bijker, W.E. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press, Cambridge, MA, 1997.

Bijker, W. and Law, J., Eds. *Shaping Technology/Building Society: Studies in Sociotechnical Change*. MIT Press, Cambridge, MA, 1994.

Bijker, W., Hughes, T and Pinch, T. Eds. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press, Cambridge, MA, 1989.

Collins, H.M. and Pinch, T. *The Golem at Large: What You Should Know about Technology*. Cambridge University Press, New York, 1998.

Ellul, J. *La Technique ou l'Enjeu du Siecle*. A. Colin, Paris, 1954. Translated by Wilkinson, J. for the English version: *The Technological Society*. Knopf, New York, 1964.

Elster, J. *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge University Press, New York, 1983.

MacKenzie, D.A. and Wajcman, J., Eds. *The Social Shaping of Technology*, 2nd edn. Open University Press, Buckingham, 1999.

Mumford, L. *Technics and Civilization*. Harcourt Brace, New York, 1934.

Mumford, L. *The Myth of the Machine, vol. 1: Technics and Human Development; vol. 2: The Pentagon of Power*. Harcourt Brace Jovanovich, New York, 1967 and 1970.

Schumpter, J.A. *Capitalism, Socialism, and Democracy*, 6th edn. Unwin, London, 1987.

Smith, M.R and Marx, L., Eds. *Does Technology Drive History? The Dilemma of Technological Determinism*. MIT Press, Cambridge, MA, 1994. Twelve essays from a conference on the theme. Bruce Bimber's "Three Faces of Technological Determinism" provides an insightful analysis.

Wajcman, J. *Feminism Confronts Technology*. Pennsylvania State University Press, University Park, PA, 1991.

Winner, L. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Theory*. MIT Press, Cambridge, MA, 1977.

Winner, L. *The Whale and the Reactor*. University of Chicago Press, Chicago, 1984.

Ziman, J., Ed. *Technological Innovation as an Evolutionary Process*. Cambridge University Press, New York, 2000.

Software Application Programs

At the beginning of the computer age around the late 1940s, inventors of the intelligent machine were not thinking about applications software, or any software other than that needed to run the bare machine to do mathematical calculating. It was only when Maurice Wilkes' young protégé David Williams crafted a tidy set of initial orders for the EDSAC, an early programmable digital computer, that users could string together standard subroutines to a program and have the execution jump between them. This was the beginning of software as we know it—something that runs on a machine other than an operating system to make it do anything desired. "Applications" are software other than system programs that run the actual hardware. Manufacturers always had this software, and as the 1950s progressed they would "bundle" applications with hardware to make expensive computers more attractive. Some programming departments were even placed in the marketing departments.

The manufacturers could not create an application for every need. Programming a computer was still an expensive and largely mysterious art. A group of International Business Machine (IBM) users, many of whom were competitors, gathered in a Los Angeles hotel room to form SHARE in 1955. This group, as indicated by its name, was a way of sharing expensive applications programs among its mostly aerospace members.

At first, most applications were created for the government, which could afford them. Business would make do with the applications included with the machines and they often had to change their business practices to match the programs. In the U.S., a large custom application was the semiautomatic ground environment, or SAGE, an air defense system. Making this application gobbled up 700 of the roughly 2000 programmers in the U.S. After it was completed, a company was formed from the remnants. This System Development Corporation kept making applications for the government, but it established a separate software industry. Computer manufacturers still threw in the software for free, but the principle of independent construction was made and the number of independent contractors and user groups therefore increased.

Also around the late 1950s, higher-level programming languages appeared. FORTRAN (FORmula TRANslator) and COBOL (Common Business Oriented Language) came into use and reduced the cost and variety of applications for government and civilians alike. Both were quickly adopted, and are alive today.

In the 1960s, there were a number of big projects using all levels of languages, such as air traffic control, the airline reservation system SABRE, and the space program. There was also a continuation of bundling. Toward the end of the decade, the government pressured computer manufacturers to unbundle software. This also made economic sense, as software was not getting cheap, but hardware was. IBM, the largest manufacturer, led the way by offering its CICS successful business product. Others soon followed suit.

After a few years, the various suppliers of application software realized that certain domains asked for the same collections of programs. They hit on the idea of supplying "packages" of relevant programs. This mirrored the sorts of routines supplied as part of bundling. This became a big industry, but custom programs were bigger. In 1970, $70 million of packages were sold, while $650 million of consulting was done.

All of these applications were primarily on mainframe or minicomputers (in other words, large ones), whereas today we are surrounded by microcomputers. The main reason for this is the development of a series of "killer apps" or highly successful and useful programs, that ran on personal computers. The first killer app was Visicalc, developed by Daniel Bricklin. This was a spreadsheet program that allowed managers to do calculations and reports on their desktop that formerly required an expensive machine and a programmer to obtain.

Along with the development of microcomputer hardware, an operating system needed to be developed to run the machine. The Microsoft Cooperation supplied the DOS (disk operating system) for the IBM PCs (personal computers), later the industry standard. Later, Microsoft developed a family of individual applications that ran both on DOS and on the Apple Computer Corporation's MacOS for the Macintosh series. Applications companies that had developed software for other operating systems ported (translated) their applications to DOS, or died. These included a word processor, spreadsheet, presentation software, and a simple database application. Some of these were combined to yield Microsoft Office, a package for microcomputers.

At the end of the twentieth century the application software with the highest sales was Microsoft Office, the database from Oracle, and the enterprise package SAP. Application software is simply what makes computers useful, and the computer revolution would not have existed without it.

See also **Software Engineering; Systems Programs**

JAMES E. TOMAYKO

Further Reading

Aspray, W. and Campbell-Kelly, M. *Computer: A History of the Information Machine*. Basic Books, New York, 1996.

Ceruzzi, P. *A History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Software Engineering

Software engineering aims to develop the programs that allow digital computers to do useful work in a systematic, disciplined manner that produces high-quality software on time and on budget. As computers have spread throughout industrialized societies, software has become a multibillion dollar industry. Both the users and developers of software

depend a great deal on the effectiveness of the development process.

Software is a concept that didn't even pertain to the first electronic digital computers. They were "programmed" through switches and patch cables that physically altered the electrical pathways of the machine. It was not until the Manchester Mark I, the first operational stored-program electronic digital computer, was developed in 1948 at the University of Manchester in England that configuring the machine to solve a specific problem became a matter of software rather than hardware. Subsequently, instructions were stored in memory along with data.

In the early days of software development in the 1950s and 1960s, that memory was expensive and precious. This inspired programming tricks that could reduce memory usage as well as increase efficiency of program execution, although there was frequently a tradeoff. Programming was labor intensive, especially when done in the binary "machine language" native to the hardware or in "assembly language" that substituted alphabetic symbols for the binary machine codes. Systematic development was not a priority as programmers focused on the essentials of making programs small enough and fast enough. English-like high-level "procedural" programming languages such as FORTRAN (FORmula Translator, designed for scientific problems) and COBOL (Common Business Oriented Language) designed for business tasks) helped make software more understandable and programmers more productive, but even then the imperatives of size and speed often remained.

Those imperatives began to change in the 1960s. Computer hardware costs began steadily decreasing while software costs (and hardware capabilities) steadily increased. Software soon became the dominant cost component of an information system. Worse yet, software was also increasingly plagued by schedule slippages and quality issues including functional defects and poor usability. These problems were epitomized by IBM's OS/360 operating system, which was released in 1966 considerably late, seriously over budget, and full of flaws. The mounting problems prompted both a label—"the software crisis"—and a response, a conference on what was provocatively termed "software engineering," sponsored by the North Atlantic Treaty Organization (NATO) Scientific Committee at Garmisch, Germany in 1968.

Participants from academia, industry, and government, traced the software crisis to several key characteristics of software as well as a general lack of discipline among programmers. First, software of any significant size is highly complex, thereby straining human cognitive capacities. Second, software is easily changed, owing to the fact that it is notation rather than a physical artifact. Poorly conceived and implemented changes can degrade quality over time. Finally, software is used to solve problems and perform work in an incredibly wide variety of areas, which makes it difficult to generalize techniques and tools as well as the software itself so as to reuse it.

Several approaches were soon proposed to address these various problems. To impose more discipline on programmers and counterbalance software's malleability, development would be guided by a life-cycle model defining each stage of the process—requirements specification, design, coding and implementation, verification (of correct implementation of the specification, usually through testing) and validation (that the software meets the user's needs), and maintenance (correction, adaptation, and enhancement). Hierarchical decomposition of programs into functionally independent modules (stepwise refinement) that hid their implementation details from other modules (information hiding) represented systematic methods for coping with complexity. Modularity and information hiding ultimately found their fullest expression in object-oriented programming, in which data and operations are bundled into "objects" that model the problem elements and whose interactions are strictly controlled. Structured programming aimed to render programs more intellectually manageable and amenable to rigorous analysis through the exclusive use of well-understood constructs of sequence, iteration, and selection. Structured programming (coding) was soon accompanied by structured (requirements) analysis and structured design. Various programming languages such as Pascal in the early 1970s, and Ada, C, and C++ in the mid-1980s emerged to better support these techniques, as did various types of computer-aided software engineering (CASE) tools.

Many of the proposed techniques, though important and useful, were perceived as lacking a certain depth. It often seemed difficult to get more specific than principles of general problem solving. Things like hierarchical decomposition (essentially divide and conquer) and abstraction (the hiding of unnecessary detail), though vital to the development of software, contrasted with the underpinnings of established scientific and technical disciplines. This concern reflected the fact that attempts to establish a discipline for software development (just like other fields) were as much

about social and professional status as about practical necessity. Furthermore, individuals with a mathematical or scientific background often had visions of a discipline very different from the ideas of those with an engineering or other type of background.

Nowhere was this tension more apparent than in arguments over formal verification of software, which sought to prove mathematically that a program satisfied its specified requirements. Formal verification was the most contentious manifestation of a broader formal methods movement that sought to apply mathematical notations and techniques, the ultimate in rigor, to virtually all aspects of software development. Those who took a more scientific or mathematically oriented view of software development argued that not only would formal verification solve software quality problems, it would make software development superior to other technical endeavors. Others strongly disputed both the feasibility and the usefulness of formal verification, given its laborious nature and dependence on correctly specified requirements.

The argument over formal verification (and formal methods generally) was emblematic of what Fred Brooks, Jr., manager of the OS/360 project, dubbed the "silver bullet syndrome." Brooks and others perceived a tendency among software technologists to seize upon a single technique or approach as the solution to the software crisis; but many of the same characteristics that engendered the software crisis in the first place also made a single, comprehensive solution unlikely.

At the end of the twentieth century, increasingly powerful software continued to be developed and used, despite the persistent problems that were once deemed a crisis. A greater appreciation for variety of technique can be seen in every phase of every software life-cycle model; and the professional status of software engineers was still debated. Software engineering's unique flavor reflected the mixtures of knowledge and practitioners that converged behind the creation of operational technical artifacts consisting only of notation.

See also **Computer Science; Software Application Programs; Systems Programs**

STUART S. SHAPIRO

Further Reading

Bergin, T. and Gibson, R., Eds. *History of Programming Languages, vol. 2*. Addison-Wesley/Longman, Boston, 1996. The product of a conference on the topic, with chapters covering significant modern programming languages.

Brooks, Jr., Frederick. No silver bullet: essence and accidents of software engineering. *Computer*, 20,4, 10–19, 1987. Insights on the nature and future of software engineering.

Brooks, Jr., Frederick. *The Mythical Man-Month: Essays on Software Engineering*, 2nd edn. Addison-Wesley Longman, Boston, 1995. Lessons from OS/360 and other reflections on software engineering.

Campbell-Kelly, M. Programming the EDSAC: early programming activity at the University of Cambridge. *IEEE Annals. Hist. Comp.*, 20, 4, 46–67, 1998. An analysis of the development of important early programming techniques.

Ceruzzi, P. Crossing the divide: architectural issues and the emergence of the stored program computer, 1935–1955. *IEEE Ann. Hist. Comp.*, 19, 1, 5–12, 1997. An analysis of the transition from programming in hardware to programming in software.

Ensmenger, N. The "question of professionalism" in the computer fields. *IEEE Ann. Hist. Comp.*, 23, 4, 56–74, 2001. An analysis of professionalization efforts in computing in the US in the 1950s and 1960s.

Pour, G., Griss, M. and Lutz, M. The push to make software engineering respectable. *Computer*, 33, 5, 35–43, 2000. An assessment of the professional state of software engineering at the turn of the century.

Sammet, J. *Programming Languages: History and Fundamentals*. Prentice Hall, Englewood Cliffs, NJ, 1969. A discussion of the origins and characteristics of early programming languages.

Shapiro, S. Boundaries and quandaries: establishing a professional context for IT. *Inf. Technol. People*, 7,1, 48–68, 1994. An analysis of professionalization efforts in computing in the U.S. and the U.K. focusing primarily on the 1970s and 1980s.

Shapiro, S. Splitting the difference: the historical necessity of synthesis in software engineering," *IEEE Ann. Hist. Comp.*, 19, 1, 20–54, 1997. An overview and analysis of the technical development of software engineering.

Solar Power Generation

The emergence of solar power generation is part of the overall movement toward renewable energy production. Interest in this type of energy production grew in the early 1970s with an increased public awareness of the negative impact of technological developments on the environment. The use of solar power, of course, was not new. Heat produced by the sun was used for all sorts of purposes from the early history of humankind. In the search for renewable energy sources, the direct use of the sun's heat has continued in the use of solar panels. In these panels, heat from the sun is absorbed by water flowing in pipes, and the hot water can then be used for heating purposes. In the twentieth century, two types of thermal solar energy systems developed: (1) active systems that

used pumps or fans to transport the heat; and (2) passive systems that use natural heat transfer processes. In 1948 a school in Tucson, Arizona, with a passive solar energy system was built by Arthur Brown. In 1976 the Aspen-Pitkin County airport was opened as the first large commercial building in the U.S. that used a passive solar energy system for heating. However, the original idea of using passive solar energy goes back to ancient times. Archeologists have found houses with passive solar energy systems dating back to the fifth century AD (see Buildings, Designs for Energy Conservation).

What was new in the renewable energy trend of the twentieth century was the conversion of solar energy into electricity in order to replace the energy that was produced from fossil, or nonrenewable sources. The device that was developed to realize this conversion is the photovoltaic (PV) cell. This cell is based on the PV effect of a number of semiconducting materials, first discovered in selenium by Willoughby Smith in 1873. In 1877 William G. Adams and Richard E. Day discovered this effect in a selenium–platinum junction, and went on to build the first selenium solar cell. The effect was subsequently seen in a variety of other semiconducting materials such as germanium and silicon. In 1954 Bell Laboratories researchers demonstrated their first solar cells, primarily for space applications. The following year Western Electric sold the first licenses for producing silicon PV cells. Commercial production of PV cells started in the same year by Hoffman Electronics Semiconductor Division. In 1958 the satellite Vanguard I was the first to be powered by PV solar cells.

Silicon is still often used for producing PV cells. The most important types of silicon cells are monocrystalline (based on single crystals), polycrystalline (based on numerous grains of monocrystals), and amorphous silicon (no crystals but thin homogeneous layers). The monocrystalline cells have the highest efficiency, but the polycrystalline cells are cheaper. Amorphous cells are the cheapest and also the thinnest type, which has advantages when the cell is to be integrated into a device. Apart from silicon, cadmium telluride and copper indium diselenide are used for making thin film cells. For thick films, gallium arsenide is an alternative material for silicon. Production processes are different for different types. Crystalline cells are produced in wafers, and amorphous cells are made by depositing the silicon on a substrate (a steel or a glass sheet covered with a layer of tin oxide).

The earliest PV cells had efficiencies of just a few percent. For example, Hoffman Electronics first cells in commercial production had an efficiency of only 2 percent, and by 1957 this had increased to 8 percent. In the course of the second half of the twentieth century, considerable research and development were done to improve the efficiency of the PV cells and to reduce the price of PV electricity. This was not without success. The average efficiency of the monocrystalline and polycrystalline silicon cells increased from 11 percent in 1985 to 16 percent in 1995. The efficiency of the amorphous silicon cells increased from 5 to 10 percent in the same period. The price of all types of silicon cells dropped to less than half the 1985 price in these years. As a result, PV electricity can now be produced for \$0.25 to 0.40 per kilowatt-hour (kWh), but this is still five times as much as electricity produced by burning coal and gas.

The structure of a PV cell is shown in Figure 12. There are several layers in the cell, and in the middle there are two semiconducting layers, one *n*-type (negative) and one *p*-type (positive). (See the entry on Semiconductors for further information on the functioning of *n*- and *p*-type semiconductors). When sunlight hits the cell, electrons in the semiconducting layers receive energy that makes them free to move. An electric field in the semiconducting layers forces them to move; and when a load is connected to the cell, an electric current can flow. This current is the PV electricity that is produced by the cell. There are two conducting layers on both sides of the pair of semiconducting layers for connecting the load. To protect the cell, there is a covering glass layer. An

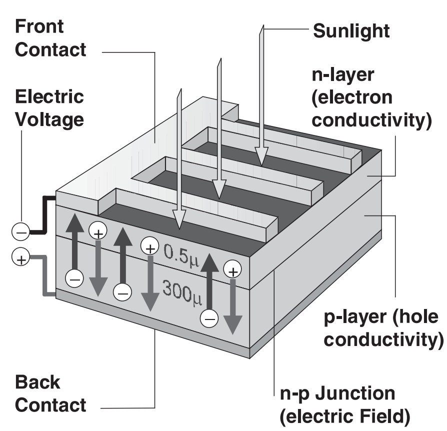


Figure 12. Diagram of a photovoltaic cell.
[*Source: From the EU Report 20015: Photovoltaics, 2001. Reprinted with permission from the Joint Research Center, Ispra, Italy*.]

antireflection layer prevents incoming sunlight from being reflected away from the semiconducting layers. The size of such cells varies from 1 to 10 centimeters in diameter.

Individual PV cells can only serve as an energy source for low-power applications. Several cells, however, can be connected and used together to form a module, also referred to as a ''solar panel,'' not to be confused with the water-based solar panels mentioned above. The PV cells produce direct current (DC). As most electric appliances are based on the use of the electricity network with its alternating current (AC), PV systems usually have a converter that transforms DC into AC. The output of an average module is 12 volts, which is converted into the 110 or 230 volts that most electric appliances need. The power that is generated by an average module is around 50 to 80 watts. The average electricity demand of a household is 1.5 to 2 kilowatts, so it is common to have around 20 to 30 modules in a PV system for supplying electricity in a house. This requires an area of around 15 by 15 square meters.

The efficiency of a complete PV energy system not only depends on the quality of the cells in the panels but also on the extent to which the changing position of the sun can be taken into account. In a number of applications the position of the panels is fixed; for example, when they are integrated into the roof of a house; in which case the house can be oriented for the optimal use of sunlight. In other cases, when a tracking system is used, the position of the panels can change, as when solar panels are used in a satellite. Tracking systems can have one or more axes in order to capture the optimum amount of sunlight. Concentrators (Fresnel lenses with concentration ratios of 10 to 500 times, or mirrors) for focusing the sunlight allow the panel to use the available sunlight more effectively. Such concentrating systems started to be used in the late 1970s.

Two types of PV cell applications can be distinguished: stand-alone and grid-connected. In stand-alone applications, the PV system functions independently of the electricity network. This type of application is usually found where connection with the network is problematic because of large distances. As there is no backup energy source in this case, a battery has to be part of the PV system. Some examples of practical applications are: energy supply for villages in developing countries; energy supply for water pumping systems; lighting of beacons in the sea; and energy supply in satellites and electrically driven boats and cars. In the case of a grid-connected system, there is an exchange of energy between the PV system and the grid. In case there is a lack of energy in the PV system (at night or dark sky days), energy can be retrieved from the electricity network. When there is a surplus of energy in the PV system, however, this surplus can be sold to the energy company by feeding it into the grid. Figure 13 is a schematic drawing of a grid-connected PV panel application in a house.

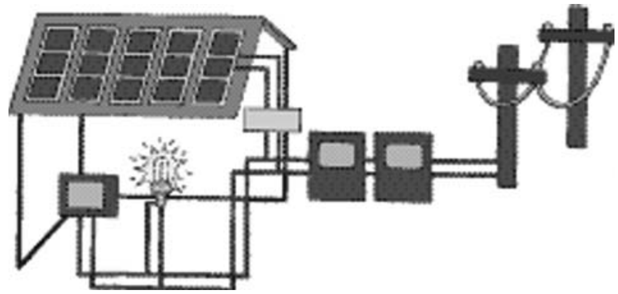


Figure 13. Grid-connected photovoltaic panel application in a house. Between the grid and the house are two meters for incoming and for delivered electricity. The converter is shown under the roof panels (right), and two electricity-using devices are shown in the house (left).

The total amount of solar energy as a contribution to the total energy production in most countries at the end of the twentieth century is still relatively small, even though it has been calculated that the sales of PV systems have increased from \$2 million in 1975 to more than \$750 million in 1993. For the year 1996 it was estimated that of the off-grid residential PV systems operational worldwide, about 10,000 were in remote vacation homes in Scandinavia. Among the reasons for the relatively low numbers is that the price of solar cells and the often-needed batteries is still too high to make a PV system economically competitive with nonrenewable (fossil fuel) energy production. A second, but less important, reason is that it is not yet clear if some types of solar cells really have better environmental properties over the whole life-cycle of the systems.

See also **Buildings: Designs for Energy Conservation; Electricity Generation and the Environment; Photosensitive Detectors; Semiconductors; Technology, Society, and the Environment**

MARC J. DE VRIES

Further Reading

Markvart, T., Ed. *Solar Electricity*. Wiley, Chichester, 2000.

Partain, L.D., Ed. *Solar Cells and their Applications*. Wiley-Interscience, Chichester, 1995.

Zweibel, K. *Harnessing Solar Power: The Photovoltaics Challenge*. Plenum Press, New York, 1990.

Solvents

Solvents are the hidden element in a broad range of technological activities, including chemical processes, paint, dry cleaning, and metal degreasing. They are used to dissolve organic compounds (water is the usual solvent for inorganic compounds) to enable reactions or polymerization, spreading or ease of use, or to extract compounds from a matrix such as plant material. Dry cleaning is a specialized form of the last group, as it removes fatty substances that adhere dirt to clothing. Many organic compounds either react with or do not dissolve in water, hence the need to find suitable organic chemicals to act as the solvent. It is hard to find compounds that dissolve a wide range of substances, but are also relatively cheap, nonflammable and nontoxic. In practice, the solvents used represent a compromise, often an unsatisfactory one.

The first solvent to be readily available was ethanol (ethyl alcohol) made by fermentation and distillation since the late Middle Ages. Amyl alcohol (fusel oil), a byproduct of ethanol manufacture, also became a popular solvent. Oil of turpentine, made from pine resin, became an important solvent in the eighteenth century and was used in paint and varnishes, and to dissolve rubber. It was also the basis of the earliest form of dry cleaning, which started around 1825. Wood spirit, made by dry distilling wood, first appeared in the early nineteenth century, but its purification into methanol (wood alcohol, methyl alcohol), acetone, and methyl ethyl ketone only took place in the middle of that century. As late as 1914, acetone and methanol were the only significant solvents in the U.S. (by volume), which had access to extensive virgin woodland. Synthetic methanol, made by treating carbon monoxide with hydrogen under pressure, was first made by the German chemical firm BASF in 1923 and eventually replaced the natural product.

Carbon disulfide, made by heating sulfur with charcoal, was the first synthetic solvent. From the 1840s onwards, it was used to dissolve oils, fats, waxes, and rubber. At the end of the nineteenth century, it became important as a major component of viscose rayon manufacture. Despite its value as a solvent, commercial carbon disulfide smells of rotten eggs and is also toxic.

Around 1820, the Scottish chemist Charles Macintosh replaced oil of turpentine by coal-tar naphtha (obtained from the growing coal-gas industry) as a rubber solvent and successfully used this solution to make the waterproof overcoats that bear his name. The real breakthrough, however, was the separation of benzene from coal-tar, pioneered by Charles B. Mansfield in the late 1840s. Benzene is an excellent solvent and available in large quantities from coal-tar. Although Mansfield died from burns caused by his distillation apparatus in 1855, benzene's flammability was tolerated and its toxicity was not considered to be a problem until the 1960s. Somewhat later in the nineteenth century, petroleum naphtha became available and competed with benzene in the dry-cleaning market. The less flammable "white spirits" fraction of petroleum became popular in the paint industry as a substitute for oil of turpentine (hence its popular if misleading name "turps") after 1900. In the mid-1920s, a variant of white spirits, Stoddard solvent (named after W.J. Stoddard, president of the U.S. National Institute of Dry Cleaning in 1928), was introduced as a less flammable version of petroleum naphtha, particularly in the dry-cleaning industry. It is still used in dry cleaning in the U.S. and Australia.

Stoddard solvent competed with a wholly different class of solvents—the chlorinated hydrocarbons. The first widely used member of the group, chloroform, was discovered in 1831. It was made cheaply from ethanol and chlorine, but its role as a solvent was soon overshadowed by the discovery of its anesthetic properties in 1847. The German chemist Hermann Kolbe made carbon tetrachloride in 1843 by reacting carbon disulfide with chlorine. First produced commercially by Chemische Fabrik Rheinau of Mannheim in 1892, it has much the same solvent properties as carbon disulfide and was also used in dry cleaning, especially for removing stains. As well as being toxic, it is also rather unstable, producing corrosive hydrochloric acid. However, the chlorinated solvents industry really took off in the early 1900s, when cheap chlorine became available as a byproduct of electrolytic caustic soda production. By 1913, the Bavarian firm Alexander Wacker, and its British associate Weston Chemicals (partly owned by the Castner-Kellner Company, which became part of ICI), were already marketing tetrachloroethane, trichloroethylene and perchloroethylene. Tetrachloroethane ("tetra" or Westron) was taken up very quickly by the dry-cleaning industry as it was an excellent solvent for fats. It was soon replaced by trichloroethylene ("tri" or Westrosol), which is more resistant to hydrolysis and less toxic. Trichloroethylene also found an important niche in the 1920s as a degreasing agent for metals. Perchloroethylene slowly became the most important dry-cleaning solvent, having beaten off stiff

competition from the chlorofluorocarbons R-11 (trichlorofluoroethane) and R-113 (trichlorotrifluoroethane) in the 1960s and 1970s, but which are now banned due to their ozone-destroying properties.

The rapid growth of car production after 1900 led to a great increase in the use of nitrocellulose-based lacquers. Initially the solvents used were based on amyl alcohol, which was expensive and in short supply. Auguste Fernbach and Chaim Weizmann developed a fermentation process for the production of butanol and acetone with the aim of making synthetic rubber. Initially, the attempt to make synthetic rubber having failed, it was the acetone that was the more valuable component, as it was used to make explosives and canvas lacquer ("dope") used to smooth and waterproof aircraft canvas frames in the World War I. The Commercial Solvents Corporation in the U.S. operated the process in the 1920s to obtain butanol for the car lacquer market. In Germany, BASF developed a route to butanol from coal-based acetylene. Ethyl acetate, produced by reacting acetic acid with ethanol, also grew in importance in the 1920s although it had been available since the mid-nineteenth century.

The development of the petrochemical industry in the 1920s and 1930s, mainly in the U.S., led to the introduction of new solvents and new routes to already established solvents such as acetone. Alcohols such as ethanol, isopropanol and isobutanol could be made by treating refinery gases with sulfuric acid. The key intermediate, ethylene oxide (initially made by treating ethylene with chlorine water and more recently by direct oxidation of ethylene), could be converted into ethylene glycol (which was a solvent for dyes and inks as well as an antifreeze) and diethylene glycol. Several exotic solvents appeared in the 1960s. The oldest member of this group, tetrahydrofuran (THF), is an important solvent for plastics. It is made from butanediol using Reppe chemistry but it can also be derived from maize, thus is capable of being a renewable chemical (see Green Chemistry). Dimethylformamide (DMF) is a powerful solvent that can be used for inorganic salts as well as plastics and pigments. Dimethyl sulfoxide (DMSO) is extremely efficient at extracting substances, and is used as a paint remover and a medium for reactions.

By the 1950s concerns were growing about the safety of several important solvents, notably benzene, carbon tetrachloride and carbon disulfide. The British factory inspector Ethel Browning was a pioneer in this field, publishing *Toxic Solvents* in 1953. Other halogenated solvents were later implicated in destruction of the ozone layer. At first, the harmful solvents were replaced by safer alternatives (benzene by toluene, carbon tetrachloride by trichloroethane and then trichloroethylene), but efforts are now focused on the replacement of organic solvents altogether. This can be achieved by using water-based emulsions (a method used in polymerization processes since the 1920s) and wherever possible, eliminating the solvent altogether ("green" chemistry). The use of ethylene dibromide (which is both carcinogenic and ozone-destroying) to extract caffeine from coffee beans has been largely displaced by a new technique, supercritical fluid extraction, which uses supercritical liquid carbon dioxide under pressure. This technique has also been adapted for dry cleaning, but the equipment is expensive.

See also **Cleaning, Chemicals and Vacuum Cleaners; Coatings, Pigments, and Paints; Green Chemistry**

PETER MORRIS

Further Reading

Cammidge, E.M. Benzene and turpentine: the pre-history of drycleaning. *Ambix*, 31, 79–84, 1991.

Durrans, T.H. *Solvents*, 6th edn. Chapman and Hall, London, 1950.

Wacker Chemie. *Im Wandel Gewaschsen: Der Weg der Wacker-Chemie, 1914–1964*. Verlag für Wirtschaftspublizistik, Wiesbaden, 1964.

Sonar

The word "sonar" originated in the U.S. Navy during World War II as an acronym for "SOund NAvigation and Ranging," which referred to the systematic use of sound waves, transmitted or reflected, to determine water depths as well as to detect and locate submerged objects. Until it adopted that term in 1963, the British Admiralty had used "ASDIC," an abbreviation for the Anti-Submarine Detection Investigation Committee that led the effort among British, French, and American scientists during World War I to locate submarines and icebergs using acoustic echoes. American shipbuilder Lewis Nixon invented the first sonar-type device in 1906. Physicist Karl Alexander Behm in Kiel, Germany, disturbed by the *Titanic* disaster of April 1912, invented an echo depth sounder for iceberg detection in July 1913. Although developed and improved primarily for military purposes in World War I, sonar devices became useful in such fields as oceanography and medical practice (e.g., ultrasound).

The earliest sonar devices used in military operations were called ''passive'' systems because they did not transmit signals. One such instrument was the hydrophone, essentially a submerged microphone hanging off the side of a ship. Developed in 1915 by French physicist Paul Langévin and Constantin Chilowsky, a Russian living in Switzerland, the hydrophone used the piezoelectric properties of quartz for powerful ultrasonic echosounding to detect submarines. The hydrophone's transducer consisted of a mosaic of thin quartz crystals glued between two steel plates with a resonant frequency of 150 kilohertz. On April 23 1916, the German submarine UC-3 became the first ship confirmed sunk after detection by hydrophone. Hydrophone efficiency improved significantly when engineers realized a pair of highly directional microphones, separated by a few feet on a connecting bar that could be rotated at the center, yielded a more accurate bearing on the sound source. Langévin's invention formed the basis of the development of naval pulse-echo sonar.

By 1918, both Britain and the U.S. had produced ''active'' systems that transmitted and received sound waves. Those systems owed much to the ingenuity of individuals seeking to improve iceberg detection. English meteorologist Lewis Richardson filed the first patent for an underwater, echoranging design at the British Patent Office shortly after the sinking of the *Titanic*. Reginald Fessenden, a Canadian living in the U.S., developed the workable ''active'' sonar, the Fathometer, in 1914. His system, which employed an electromagnetic moving-coil oscillator that emitted a low-frequency signal, and then switched to a receiver to listen for echoes, detected an iceberg underwater at a distance of more than 3 kilometers.

Although the development of sonar came too late to have any significant effect on the course of World War I, its improvement and widespread shipboard installation during the interwar years profoundly influenced naval operations during World War II. In 1921, a sonar system installed aboard the *HMS Antrim* could detect a shutdown submarine lying on the ocean bottom at a distance of 1800 meters. The following year, the U.S. Navy employed on the survey ship *USS Stewart* an echosounder designed by Harvey Hayes. Many French ocean liners, beginning with the *Ile de France* in 1928, carried Langévin's echosounding devices. Meanwhile, Rudolf Kühnhold at the German navy's Nachrichtenmittel-Versuchsanstalt (NVA) in Kiel began devising a sum-difference method of sound location for directing gunfire at surface or underwater targets and, by the early 1930s, developed a close working relationship with Tonographie company founders Paul-Günther Erbslöh and Hans-Karl Freiherr von Willisen. They subsequently created a spin-off firm, GEMA, to produce the sophisticated electronic sounding equipment needed by the German military. In 1931, the U.S. Navy Underwater Sound Group produced the ''QB'' echoranging sonar, which was effective below 6 knots, and the first installation of echoranging equipment on American destroyers occurred in 1934. By the beginning of World War II, over 200 British warships carried sonar equipment that was effective up to 15 knots. Although sonar was responsible for approximately 60 percent of all submarine kills during the first two years of World War II, it proved to be an ineffective tool for sweeping broad areas.

Sonar improvements during World War II enabled ever more precise detection, tracking, and targeting of ever more capable, deeper-diving submarines. In 1942, the ''Q'' attachment permitted tracking of U-boats at closer ranges than the earlier ''searchlight'' sonar allowed. Before the end of 1943, introduction of the Type-147 or ''Sword'' system added the capability to track deep-diving U-boats laterally and thereby to make last-second targeting adjustments. Meanwhile, American submarines employed the bathythermograph to detect thermoclines, layers of water where the temperature gradient is greater than that of the warmer layer above and the colder layer below, beneath which they could escape sonar detection by German or Japanese surface ships. Sonar also became useful in guiding depth charges and torpedoes to their targets, as well as in detecting subsurface mines.

The Cold War that characterized the latter half of the twentieth century brought advances in sonar technology. Cognizant of the success of German U-boats during World War II and concerned about Soviet expansionism, the U.S. Navy relied for many years on the sound surveillance system (SOSUS), an essentially passive, worldwide network of underwater microphones, to detect enemy submarines. Helicopters, specially equipped with both passive and active sonar gear, also became important platforms in antisubmarine warfare.

During the 1980s, the U.S. and other naval powers began experimenting with active, long-range sonar systems that used high-intensity, low-frequency sound waves. Approximately 39 ships participated in the low-frequency active (LFA) sonar testing, which proved an LFA-equipped vessel could detect an enemy ship hundreds of kilometers distant. In March 1995, the U.S. Navy also began testing aboard the *USS Asheville* a high-

frequency system that enabled the submarine to navigate more effectively in shallow water. All the high-intensity, active sonar systems drew protests, petitions for action, and lawsuits from conservation and animal welfare organizations, because they caused widespread harm to marine mammals and other ocean life. In October 2003, the U.S. Navy agreed to restrict peacetime use of its surveillance towed array sensor system (SURTASS) LFA system, leaving environmentalists to seek similar limitations on mid- and high-frequency sonar.

See also **Submarines, Military; Submersibles; Ultrasonography in Medicine**

RICK W. STURDEVANT

Further Reading

Gannon, R. *Hellions of the Deep: The Development of American Torpedoes in World War II*. Pennsylvania State University Press, University Park, PA, 1996.

General Accounting Office. Report to the Acting Secretary of the Navy; Undersea Surveillance: Navy Continues to Build Ships Designed for Soviet Threat (GAO/NSIAD-93–53). General Accounting Office, Washington D.C., 1992.

Gerken, L. *ASW Versus Submarine Technology Battle*. American Scientific Corporation, Chula Vista, California, 1987.

Hill, J.R. *Anti-Submarine Warfare*. Naval Institute Press, Annapolis, MD, 1993.

von Kroge, H. *GEMA: Birthplace of German Radar and Sonar*. Institute of Physics Publishing, Bristol, 2000.

Urick, R.J. *Principles of Underwater Sound*. McGraw-Hill, New York, 1983.

Useful Websites

Bellis, M. *The History of Sonar*: http://inventors.about.com/library/inventors/blsonar.htm

Department of the Navy, U.S. *SURTASS LFA*: http://www.surtass-lfa-eis.com/index.htm

Derencin, R. *Counter Sonar Measures—A Bathythermograph*: http://uboat.net/articles/?article=45, http://uboat.net/articles/?article=45

Derencin, R. *Underwater Sound Detector in WW1 and WW2*: http://uboat.net/articles/index.html?article=52

Federation of American Scientists. *Introduction to SONAR*: http://www.fas.org/man/dod-101/navy/docs/es310/uw_acous/uw_acous.htm

IUSS-CAESAR Alumni Association. *The Cable*. http://www.iusscaa.org/ide.htm

McDaniel, J.T. *Hydrophones, Sonar, and Other Listening Gear*: http://www.authorsden.com/externalsite.asp?authorID=2136&destURL=www%2Efleetsubmarine%2Ecom&msg=

Moreavek, L.T., Leonard, J.G. and Brudner, T.J. *USS Asheville Leads the Way in High Frequency Sonar*: http://www.fas.org/man/dod-101/sys/ship/docs/uss_asheville.htm

Proc, J. *Asdic, Radar and IFF Systems Aboard HMCS Haida*: http://webhome.idirect.com/~jproc/sari/sarintro.html

Tanaka, T. *et al. Low Frequency Synthetic Aperture Sonar*: http://www.nec.co.jp/techrep/en/r_and_d/r03/r03-no2/rd04.pdf

USS Francis M. Robinson Association. *Sonar Equipment*: http://www.de220.com/Electronics/Sonar/Sonar.htm

Space

Space technology has a unique place in the history of twentieth century technology, since it concerns the application of technology beyond the confines of the earth. In engineering terms, this has involved communications (since signal travel time may be measured in minutes), ruggedness (since spacecraft must survive a hostile thermal and radiation environment), and reliability (since spacecraft beyond low earth orbit cannot be retrieved or repaired).

Space technology is also unusual because it is both a user of other technologies—electrical, electronic, mechanical, power, computing, and telecommunications—and a developer of these technologies. The challenge of developing systems that operate reliably in the space environment and hardware that is sufficiently light and compact to be launched in the first place is so great that developments in space technology often led to improvements in earthbound systems. One example is the microminiaturization of electronic components required for early spacecraft, which drove the development of electronics and computer systems of the late 1950s and early 1960s.

The degree to which this ''spin-off'' occurred is difficult to prove, since it is impossible to remove the influence of space exploration and development from the history of the twentieth century. However, the subject has proved itself to be important in at least two main ways. First, space is important as a place to go: as the mountains and the oceans have to past explorers, it provides a goal and a challenge with which to satiate mankind's inherent desire to explore (summarized for many in the phrase ''the final frontier''). Second, space is important as a platform from which to provide services: from satellite communications and GPS navigation, to remote sensing of the earth and the cosmos.

Here, we look at the development of space technology in the twentieth century, from its roots in science fiction to its industrialization and later commercialization. We discuss the impact of its various applications on society and its influence on culture in the second half of the century. Although

these factors have varied in degree from nation to nation, space technology has had a global impact; indeed it has been a key factor in the development of ''globalization'' itself.

Fictional Roots

It seems likely that space travel has been part of human imagination since mankind realized that space was ''somewhere to go.'' As a planet, earth is unusual in that it has a moon that appears relatively large in its sky, and even with the naked eye, considerable detail is visible on its surface. This alone must have evoked speculation as to the possibilities of traveling there, even before the invention of the telescope.

Although thirteenth century China is credited with the invention of the rocket, Cyrano de Bergerac is believed to have been the first to propose that rockets should be used as a form of space propulsion. In his novel *The Comic History of the States and Empires of the Moon and the Sun*, published in 1649, he imagined being raised aloft by rockets attached to a flying machine.

Jules Verne, however, is credited with writing the first fictional account of spaceflight based on scientific fact: *From the Earth to the Moon*, which appeared in 1865. Since the bullet was by far the fastest thing in Victorian life, his fictional spacecraft was a bullet-shaped projectile propelled by a cannon. Verne's first astronaut crew would have been pulverized by the acceleration, but at least he had recognized the need for speed to escape the earth's gravity as described by Newton in the seventeenth century.

Once the technology of ''moving pictures'' had been developed, it was inevitable that it should be used to portray what we now call science fiction. Indeed, a number of science fiction ''shorts,'' of only a few minutes in length, were made prior to 1900. However, the first feature-length (21-minute) science fiction film was *Le Voyage dans la Lune* (A Trip to the Moon), made by George Méliès in 1902; it was inspired by Verne's *From the Earth to the Moon* and H.G. Wells' *The First Men in the Moon*. Thus began a long line of space-related films which, for some, culminated in the 1968 classic *2001: A Space Odyssey*, released the year before Neil Armstrong became the first man to set foot on the moon.

Military Funding

The early pioneers such as Konstantin Tsiolkovsky and Robert Goddard (see entry on Rocket Propulsion, Liquid Propellant) received little public support or government funding for their early work, and in common with many technologies, the development of the rocket was driven by the needs of warfare. In 1932 Wernher von Braun was employed by the German army's rocket artillery unit to develop ballistic missiles. In World War II, the result was the V-2, which later became the basis for both Russian and American rocket programs. Indeed, it was Wernher von Braun who became the leading proponent of, first, the U.S. Army's and later, the National Aeronautics and Space Administration's (NASA's) space exploration programs of the 1950s and 1960s, culminating in his work on the Saturn-V moon rocket (see Figure 14).

Even today, it is difficult to reach a consensus on so-called ''dual-use technologies,'' such as rockets and remote sensing satellites: the former can be used as spacecraft and weapon delivery systems and the latter can provide high-resolution images for both civil and military applications. Although the same dichotomy exists in the use of aero-engines, terrestrial surveillance systems and computer technology, a satellite-targeted, global positioning system- (GPS-) guided, nuclear-tipped intercontinental ballistic missile (ICBM) has far greater destructive potential.

Politics and the Space Age

It is undeniable that the early developments of the space age, which began with the launch of Sputnik-1 on 4 October 1957, were driven as much by politics as they were by science or engineering. Both the USSR and the U.S. were technically capable of launching the first satellite in the late 1950s. The reason the Soviet Union won what became known as the ''space race'' boils down to two complementary factors: a desire on the part of the Soviet leadership to prove superiority in at least one high-technology field; and an incapacity on the part of the American leadership to believe that the Soviet Union was capable of such a feat (despite the fact that the launch of Sputnik-1 had been trailed several months before it occurred).

Indeed, it was the American public (led no doubt by the media) that was most outraged, and worried, by the Soviet triumph. They believed that if the USSR could place a satellite into orbit, it could just as easily target a nuclear warhead on an American city. These feelings of vulnerability and inferiority, fuelled by further Soviet ''space firsts''—including the first moon probes, first spacecraft to Venus, first animal in space, and first man in space—led to President Kennedy's

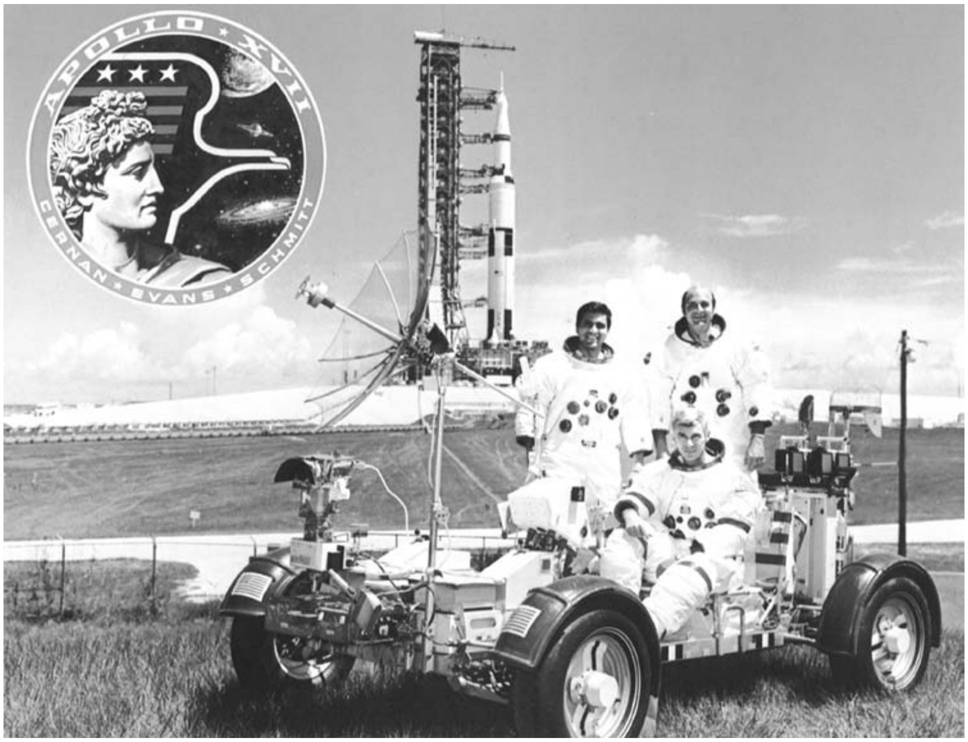


Figure 14. The Apollo-17 astronauts (Harrison Schmitt, Ron Evans and Gene Cernan) pose in their lunar roving vehicle (LRV) in front of the Saturn-V rocket that took them to the moon. Apollo-17, launched in December 1972, was the final mission in the Apollo program and the third to carry an LRV to the moon.
[*Source: Courtesy of NASA*.]

historic decision to commit his nation to landing a man on the moon before the end of the 1960s.

This political decision, and the realization of its goal in July 1969, provided a spur to the development of space technology that has never been seen since. It is significant not only in the scope of the history of technology, but of history itself, that only eight years after Yuri Gagarin became the first human being to orbit the earth and less than 12 years after Sputnik-1, two men were standing on the surface of the moon (Figure 15).

It has often been suggested since the end of the Apollo program that a similar venture, if attempted today, could not achieve the same result in the same timescale. This is probably true because no one can envisage reproducing the level of political and financial support that existed in the 1960s.

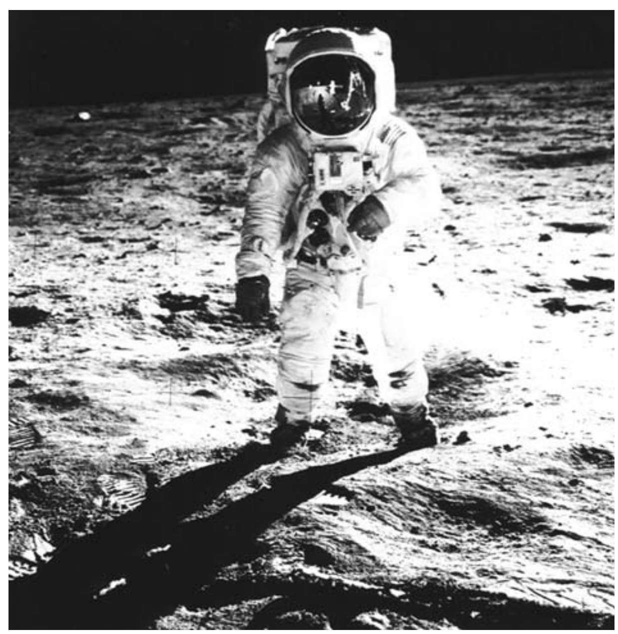


Figure 15. Edwin "Buzz" Aldrin, the second man to set foot on the moon in July 1969. Although often captioned incorrectly as a picture of Neil Armstrong, this photograph was taken by Armstrong, who appears reflected in the visor of Aldrin's helmet.
[*Source: Courtesy of NASA*.]

Industrialization

While space was a mainly scientific endeavor in the late 1950s, typified by investigations of radiation belts and solar particles, in the 1960s it became industrialized. Although government laboratories

and rocket test ranges operated by the military could handle the requirements of the early space age, the level of organization required to send manned spacecraft to the moon was something else. Indeed, apart from any technological spin-offs, the NASA-led Apollo program is credited with much of the development of modern management practice. Never before had it been necessary to coordinate so many different industrial contractors, in so many states, to produce so many leading-edge technologies on such a short timescale. When Kennedy made his Apollo commitment in May 1961, America had just 15 minutes and 22 seconds of experience in manned spaceflight, accrued on just one suborbital hop, which reached an altitude of 187 kilometers. The moon, by contrast, was about 375,000 kilometers away and a successful lunar mission would last at least eight days.

The developments required to meet Kennedy's challenge were undertaken largely by the nascent space industry, an outgrowth from the existing aerospace industry. Even after the glory days of Apollo, companies appeared proud of their space divisions. Especially for defense contractors, which were obliged to keep much of their work secret, the civilian space industry offered a "shop window" for their capabilities. Likewise, companies that were known mainly for earthbound products, such as cars and refrigerators, used their involvement in space to indicate their advanced, high-tech status. For example, it is not widely known outside the space industry that Ford, renowned for its motor vehicles, was one of America's top three satellite manufacturers until 1990, when its space division was bought by the Loral Corporation. And Bendix, which among other things manufactured washing machines, was the losing bidder (to Boeing) for the Apollo lunar roving vehicle. In the 1960s and 1970s, corporations of all different shapes and sizes wanted a piece of the action in space.

As enthusiasm for manned spaceflight waned in the U.S. (no American astronauts were launched between the Apollo–Soyuz docking mission in 1975 and the first Space Shuttle launch in 1981), the space industry polarized its efforts in two main directions: military and civilian. Already familiar with the needs of military agencies, contractors continued to develop military communications and reconnaissance satellites, while their civil divisions concentrated on similar satellites for civil government and commercial customers. Still other divisions continued to develop launch vehicles in a type of symbiotic relationship: as satellites grew larger and more complex, they needed more capable launch vehicles; and as launch vehicles grew more capable, it became possible to specify heavier satellites.

America was the undisputed leader in most aspects of space technology in the 1960s, but it was challenged later by other nations as they too developed a space industry. For example, although Russia operated under an entirely different political and industrial system, it took the lead in manned spaceflight in the 1970s and 1980s with its deployment of the Salyut and Mir space stations. Meanwhile, Europe's introduction of the Ariane launch vehicle in 1979 led to its domination of the commercial launch industry in the late 1980s and 1990s.

Several other nations developed their own space industries, predominantly based on government as opposed to commercial programs. Japan, for example, has an industry founded on science applications that has benefited from U.S. technology transfer programs for both satellites and launchers. Many of its satellites have been built jointly with U.S. contractors and its early launch vehicles were effectively American vehicles built under license by domestic contractors. By the end of the century, this had placed Japan in a position to embark on the commercialization of its satellite industry and its H-IIA launch vehicle.

India, too, is a nation that recognized the advantages of space systems. For example, the unique advantage of the communications satellite in geostationary orbit is that it can provide coverage to an entire nation, interconnecting villages that will probably never be linked by terrestrial telecommunications cables. At the same time, satellites can provide television signals to an entire nation, an advantage that has been used by India to provide education services and health advice, as well as entertainment, to millions of people who would otherwise remain deprived. By the same token, remote sensing satellites, which provide wide geographical coverage on a repeatable basis, are used for weather forecasting, flood monitoring, crop-health monitoring and many other applications in support of an ever-growing population. Despite its status as a less developed country with a large underprivileged and undereducated population, India made a political decision to develop both satellites and launch vehicles manufactured by a domestic space industry.

Global and Cultural Impact

The development of space technology has had an undeniable impact on our understanding and

appreciation of the earth, largely because it allowed us to take a global view. For example, the first pictures showing earth as a planet in space were taken from spacecraft in lunar orbit, the most famous being the "earthrise" sequence from Apollo-8 (see Figure 16). It is often suggested that these images provided the inspiration for the green movement and kick-started a new phase of global awareness; certainly they have been widely used in the print media ever since.

The communications satellite has, arguably, done as much as any other technology to make the "global village" a reality. Thanks to the satellite, it is now possible to communicate from any position on the earth's surface to any other; uplink news reports from anywhere on earth using equipment the size of a briefcase; and ascertain one's location anywhere on earth using the global position system (GPS). The satellite has also provided a tool for democracy, again because of its global coverage and its insensitivity to political boundaries. It is for this reason that, as the twentieth century closed, individual satellite receiving antennas remained illegal in China.

Apart from its economic, social, and political influences, space technology has had a widespread impact on our increasingly global culture. The evidence that space is part of our culture is there for all to see—in books, newspapers, films, and on television. Incredible though it may seem, after more than 40 years of the space age, space technology is still used in advertising to suggest technical advancement and high reliability, while children's toys still reflect a fascination with space, especially at primary level, where space is reckoned to be second only to dinosaurs in the interest ratings.

On the broadest level, space technology has enhanced our understanding of the "cosmos" and mankind's place in it, an understanding that has both intellectual and religious ramifications. We have absolute proof that the earth is neither flat nor the center of creation, and we know far more about the objects, processes, and scale of the universe than we would without space-based astronomy. From a psychological standpoint, we have proved that mankind is not necessarily limited to living on a single planet, and the development of space technology has shown that "the final frontier" is simply another boundary to cross. Perhaps the ultimate space development is the one that will enable tourists, as opposed to career astronauts and cosmonauts, to cross that boundary. The first step towards this goal was made in 2001, when Dennis Tito paid the Russian authorities a reported US$20 million for a one-week visit to the International Space Station.

As with any technology, space technology can be applied in many ways. It is possible that its future development will allow mankind to conduct wars in space and ultimately destroy the earth. It is equally possible that space technology may one day save mankind from a devastating asteroid impact, a concept that has evoked serious consideration from both film makers and government officials (Figure 17). Certainly, by the time the sun reaches its red giant phase and expands to encompass the earth, only space technology will have offered salvation.

Figure 16. "Earthrise" from lunar orbit, as photographed in December 1968 during the Apollo-8 mission. This was the first time anyone from earth had seen their home planet from the moon. [*Source: Courtesy of NASA*.]

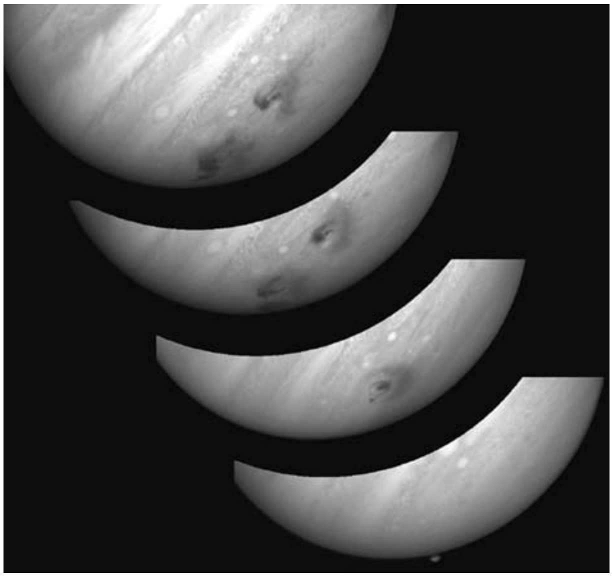


Figure 17. The impact of comet Shoemaker-Levy on Jupiter in 1994, imaged by the Hubble space telescope, showed that earth too is at risk; the circular marking on Jupiter's clouds is approximately the same size as the earth. Space technology is the only technology that offers options to either destroy or deflect such an object.
[*Source: Courtesy of NASA.*]

In the nearer term, and in an historical context, it is relevant to consider what the twentieth century will be remembered for in, say, another thousand years. Unless the planet has suffered another "dark age" and all records have been lost, it will be remembered as the century in which mankind harnessed the techniques of space technology and learned to live and work in space. As Konstantin Tsiolkovsky once said, "earth is the cradle of mankind, but one cannot remain in the cradle forever."

See also **Rocket Propulsion, Liquid Propellant; Satellites, Communications; Space Exploration, Planetary Landers; Space Exploration: Moon, Manned; Space Shuttle; Space Stations, Mir**

MARK WILLIAMSON

Further Reading

Allday, J. *Apollo in Perspective: Spaceflight Then and Now*. Institute of Physics, Bristol, 2000.

Allward, M., Ed. *The Encyclopedia of Space*. Hamlyn Publishing Group, London, 1968–1969.

Bilstein, R.E. *Orders of Magnitude: A History of the NACA and NASA, 1915–1990*. NASA, Washington D.C., 1989.

Canby, C. *A History of Rockets and Space*. Leisure Arts, London, 1964.

Cooper, H.S. F. *Moonwreck—13: The Flight That Failed*. Granada/Panther, St. Albans, 1975.

Editors of Fortune. *The Space Industry: America's Newest Giant*. Prentice Hall, Englewood Cliffs, NJ, 1962.

Furniss, T. *Manned Spaceflight Log*. Jane's, London, 1983.

Gatland, K. *The Illustrated Encyclopedia of Space Technology*. Salamander Books, London, 1981.

Godwin, R. *Rocket and Space Corporation Energia*. Apogee Books, Burlington, 2001.

Hardy, P. *The Encyclopedia of Science Fiction Movies*. Octopus Books, London, 1986.

Harvey, B. *The New Russian Space Programme: From Competition to Collaboration*. Wiley-Praxis, Chichester, 1996.

Harvey, B. *The Chinese Space Programme: From Conception to Future Capabilities*. Wiley-Praxis, Chichester, 1998.

Harvey, B. *The Japanese and Indian Space Programme: Two Roads into Space*. Springer-Praxis, Chichester, 2000.

Krige, J. and Russo, A. *Europe in Space 1960–1973*. European Space Agency, Noordwijk, 1994.

Lewis, R.S. *The Illustrated Encyclopedia of Space Exploration*. Salamander Books, London, 1983.

Lovell, J. and Kluger, J. *Lost Moon: The Perilous Voyage of Apollo-13*. Houghton Mifflin, Boston, 1994.

Miller, R. *The Dream Machines: A Pictorial History of the Spaceship in Art, Science and Literature*. Krieger, Malabar, 1993.

Murray, C. and Cox C.B. *Apollo: The Race to the Moon*. Secker & Warburg, London, 1989.

Ordway, F.I. *Visions of Spaceflight: Images from the Ordway Collection. Four Walls Eight Windows*, New York, 2001.

Ordway, F.I. and Sharpe, M.R. *The Rocket Team*. Heinemann, London, 1979.

Piszkiewicz, D. *Wernher von Braun: The Man Who Sold the Moon*. Praeger, Westport, CT, 1998.

Reeves, R. *The Superpower Space Race*. Plenum Press, New York, 1994.

Schneider, S. *Collecting the Space Race (with price guide)*. Schiffer, Atglen, 1993.

Stoiko, M. *Pioneers of Rocketry*. Hawthorn Books, New York, 1974.

Stuhlinger, E. and Ordway, F.I. *Wernher von Braun: Crusader for Space*. Krieger, Malabar, 1994.

Williamson, M. *The Cambridge Dictionary of Space Technology*. Cambridge University Press, Cambridge, 2001.

Space Exploration, Fly-Past

Flying past a target in space is one of the oldest and cheapest means of space exploration and is still the method engineers choose for first investigations in space. It returns often surprisingly good science with a simple mission profile. At the beginning of space flight, rockets were too weak, navigation too unreliable, and probes too small to do anything but simply fly by. Therefore most of the fly-pasts were in the first decade of the space age, carrying fields and particles experiments, and sometimes, primitive cameras. Later, fly-pasts were continued to reduce the cost and complexity of missions or to gain gravity boosting. Often the unknown nature

of the exploration precludes anything but a fly-past. A simple variation on the fly-past mission is the "strike" mission. Such missions were limited to the moon at first because it was thought no life could exist there, and therefore bacteria carried on a spacecraft could not contaminate indigenous life. Also a strike mission would be an early navigational *tour de force*, and the probe would be in line-of-sight of the ground stations here on earth.

Earth's moon was the first of the planets to be investigated by spacecraft. In 1959, the American Pioneer-4, with primitive instrumentation and camera, made the first flight to go near to the moon, if 60,050 kilometers is considered near for a strike mission. No pictures were recorded because the spacecraft did not come close enough to trigger the camera's photoelectric sensor. Following successful fly-past, signals from the spacecraft were received from a record 655,300 kilometers away, amazing for an early mission that was planned to go half as far. Later that year the Soviet Luna-3 swept by, carrying a primitive imaging system that returned a set of photographs of the hidden side of the moon, seen for the first time by inhabitants of the earth. Failed strike missions, Pioneer-3 and Luna-1 and *-2*, preceded both. The Soviets only sent one more intentional flyby mission. The Zond-3 of 1965, which was actually a Mars mission that had missed its launch window and was retargeted, also imaged the far side of the moon, after several failed lander missions. The U.S. sent a barrage of nine Rangers, primarily equipped with cameras, intended for impact on the moon, and several of these became inadvertent flybys when their navigation failed. Both countries quickly graduated to orbiters and landers due to their interest in the size and proximity of the moon.

While Mars and Venus held early fascination for space exploration, both planets are much smaller and farther away. Mariner-II flew past Venus in 1962 with a set of instruments including infrared and microwave sensors, radiometers, magnetometer for exploring space and the planet. The results established the magnitude of the inhospitable surface temperatures, measured at 425°C, and the absence of a magnetic field. No cameras were used because opaque clouds hide the surface of Venus. The Soviets also had several craft fly by Venus, leading to the Venera-3 strike mission (which may have been a failed lander). There were no great discoveries, either because they were too far away for their instrumentation or because of radio failure. Mariner-IV passed Mars in 1964, shooting the first close-up photos of another planet, showing Martian craters, and gathering atmospheric data. The 21 pictures took one week to transmit to earth.

The last of the Mariner missions was Mariner-X, an intentional multiple fly-past of Mercury by way of a Venus near miss to gain gravity-boost, represented the pioneering use of this technique. It carried a high-resolution camera, and the probe imaged Mercury three times, using the planet's gravity to gain the multiple fly-pasts.

During the 1970s, it became apparent that the outer planets would be lined up in such a way that a "grand tour" mission could be planned that would fly past all the gas giants—Jupiter, Saturn, Uranus and Neptune. Financially, this was deemed impossible, so a pair of Voyager spacecraft were prepared for only Jupiter and Saturn fly-pasts. Once launched, and with both probes operating, the grand tour would be deemed viable. Since instrumentation was not intended to view Uranus and Neptune, new software was radioed to the spacecraft, for example, for image compression so that more pictures could be transmitted. Also, because of lower levels of light than expected, new techniques were developed to allow the cameras to track targets during longer exposures. Voyager-2 thus visited Jupiter, Saturn, Uranus and Neptune on one of the most valuable fly-past missions ever, returning spectacular images. The Voyagers had, in addition to the high-resolution cameras, instruments for exploring areas of space such as magnetic fields. They also carried a message from earth on a disk on the spacecraft. The second probe to each of the giants, Galileo to Jupiter, and Cassini to Saturn, were orbiters carrying entry probes. Galileo's entered the Jovan atmosphere, and Cassini's is aimed at the Saturn moon Titan. Both missions depended on data gathered from *the Voyager* fly-pasts for mission planning.

The close approach of Halley's Comet in the mid-l980s spawned a plethora of fly-past missions. The Soviets had two, Vega-I and -II, both of which dropped probes into the Venusian atmosphere on their journeys. The Japanese also had a pair, Sakigake and Suisei. Only Suisei had imaging equipment. They carried instruments to determine the composition of the comet. The European Space Agency sent one probe, Giotto, which went on to another comet later after a period of dormancy. These missions were the first visits to a comet, which has a very weak gravitational field, so fly-pasts were in order for both reasons of novelty and technical limitations.

The only planet remaining to be explored, Pluto, was also targeted for a fly-past, hopefully to arrive before the atmosphere of the planet begins to

freeze as it moves farther from the sun. The earliest methods of space exploration will continue to be used, especially when exploring a new part of the universe.

See also **Space Exploration, Unmanned; Space Exploration, Planetary Landers; Space Exploration: Moon, Unmanned; Telescopes, Space**

JAMES E. TOMAYKO

Further Reading

Anderson, Jr, F.W. *Orders of Magnitude: A History of NACA and NASA*, 1915–1980. NASA, Washington D.C., 1981.

Ezell, E. *On Mars: The Exploration of the Red Planet, 1958–1978*. NASA, Washington D. C., 1984.

NASA Historical Data Books, vol. 5. NASA Launch Systems, Space Transportation, Human Spaceflight and Space Science, 1979–1988. NASA, Washington D.C.

Sheldon II, C. *Review of the Soviet Space Program*. McGraw-Hill, New York, 1968.

Siddiqi, A. *Challenge to Apollo*. NASA, Washington D.C., 2000.

Washburn, M. *Distant Encounters: The Exploration of Jupiter and Saturn*. Harcourt Brace Jovanovich, New York, 1984.

Useful Websites

NASA website: http://history.nasa.gov

Space Exploration, Manned Orbiters

Survival in space requires more than just a sturdy ship suitable for the explorers of the past. Manned orbiting vehicles provide the safe, hospitable, and transportable environment found on the surface of the earth that humans need to survive the harsh reality of outer space. From capsules to shuttles to space stations, American, Soviet, and later Chinese space pioneers all took steps toward conquering the unknowns of space in the twentieth century.

Astronauts and cosmonauts alike certainly found the first generation of spacecraft designs cramped and uncomfortable. The designs differed, but their capacities in terms of room to stretch were equally limited. The Soviet Union succeeded in launching the first human into orbit aboard the Vostok spacecraft. The spacecraft measured 7 meters long and included both a spherical re-entry vehicle and an instrument module that was jettisoned as the craft began its descent back into the atmosphere. The re-entry vehicle, which housed the cosmonaut for the duration of the orbital flight, measured only 2.3 meters in diameter. The spherical design of the re-entry vehicle allowed for even distribution of heat across its surface as it entered the atmosphere. Because the Soviets landed their craft on land instead of in the ocean as the Americans did, the cosmonaut ejected from Vostok at a safe altitude and parachuted down to the earth while Vostok pounded into the ground, only slowed by its own parachute. The Americans' first flights into orbit were aboard the Mercury capsule. Largely conical in shape, the bottom surface of Mercury was coated with an ablative heat shield for re-entry. Unlike Vostok, the entire orbiting vehicle returned to earth with the astronaut aboard; but it measured even smaller in size: just 3.3 meters long, including the retro rockets and 1.9 meters across at the heat shield's widest point.

When the two space programs moved toward the moon and began carrying more than one person at a time, the Soviets' concept changed few elements of its design. The idea of ejecting two or more cosmonauts for landing created a more challenging engineering problem than safely ejecting one. The Voskhod was the result. Cosmonauts could land aboard Voskhod once engineers added a retro-rocket system to the vehicle; but given the very tight confines of the vehicle itself, when three cosmonauts were aboard, they could not wear spacesuits during the flight, leaving the crew unprotected from the dangers of depressurization. The Gemini capsule looked much like a larger version of Mercury, at a height of 7.7 meters and weighing 3800 kilograms, two and a half times Mercury's weight. However, its modular component design and the addition of a docking mechanism made Gemini a more flexible and useful design than Mercury.

Part of the Gemini and Voskhod missions included spacewalks in preparation for moon flights. The Voskhod vehicle did not carry enough air to pressurize the entire capsule; so on March 18 1965, Aleksei Leonov used an airlock attached to the side of Voskhod for his spacewalk. With the smaller volume of the Gemini vehicle, Ed White was able to conduct his spacewalk on June 3 1965 without an airlock. James McDivitt stayed inside the Gemini 3 spacecraft wearing his own spacesuit while White floated just outside the hatch.

With the successes of Gemini—they included a spacewalk and a rendezvous with the Agena satellite—the American space program forged ahead with the flight of its newest conical spacecraft, Apollo. Similar to Voskhod, the Apollo spacecraft supported three astronauts. Wider than Voskhod, the Apollo spacecraft, measuring 3.9 meters across at its widest point, allowed the seats to be placed side-by-side. Aboard Voshkod, technicians had to raise the middle seat by several

inches to allow the cosmonauts' shoulders to overlap. Apollo also produced a second manned orbiter, the lunar lander.

The lunar landing module, spider-like in appearance, represents the flimsiest of NASA's spacecraft. Designed for the one-sixth gravity and no-atmosphere environment of the moon, the lunar lander stood on spindly legs and was covered by sharp angles. Unlike other manned space vehicles, it had a very limited purpose in terms of crew support. The crew would spend the majority of its flight to the moon and back in the Apollo command module. En route to the moon, the command module pilot turned the command module 180 degrees and docked with the lunar lander. The commander and lunar module pilot only entered the lunar lander prior to their departure for the surface of the moon. More than any other spacecraft, the lunar lander served as a dinghy; the crew used it to get to and from its sailing vessel. No part of the lunar lander returned to earth.

Although the Soviets never achieved a manned lunar landing, they too designed a vehicle for the purpose, which would eventually fly. Sergei Korolëv, the Chief Designer of the Soviet space program, and his team developed Soyuz with docking in mind. Like their previous vehicles, the basic structural design of Soyuz was a sphere. Three modules—the instrument module, the orbital module, and the descent module—connected together to give the crew a ship with both laboratory and living spaces. In January 1969, Soyuz-4 and -5 docked together and transferred crew members for the first time, an achievement that the Americans would have to perform as well for the success of the moon program. The Soviet skills and technical achievements with docking procedures contributed to their next space success, the first space station, Salyut-1, launched on April 19, 1971.

The Soyuz and Apollo spacecrafts both saw service in the 1970s in the cooperative joint mission, the Apollo-Soyuz Test Project. The program brought American and Soviet space enthusiasts together for the first time in friendship instead of competition. However, the shuttle era of the 1980s saw a less intense but seemingly renewed race. In 1983, the U.S. launched the space shuttle Columbia, the first space truck. Designed both as a rocket and a plane, the shuttle launched on the back of solid- and liquid-fuel engines, then landed like an airplane on a flat runway. In November 1988 the Soviets launched their shuttle Buran, almost identical to the American design; but the collapse of the Soviet Union in August 1991 left the Soviet space program without funds and little hope of a future. With the slow destruction of America's own fleet of space shuttles, the shape of the world's future manned space orbiter designs are again flexible. The Chinese entered the boundaries of outer space in the late 1990s. Their vehicle bore a marked resemblance to the Soyuz descent module, an igloo-like capsule. In the years since the Soviets first put a man into orbit, the vehicles going into space have come full circle and the future of manned space orbiters remains to be seen.

See also **Space; Space Exploration, Unmanned; Space Shuttle**

AMY FOSTER

Further Reading

Gatland, K.W. *Manned Spacecraft*. New York, 1976.
Jenkins, D.R. *Space Shuttle: The History of the National Space Transportation System*. Cape Canaveral, FL, 2001.
Neal, V., Lewis, C.S. and Winter, F.H. *Spaceflight: A Smithsonian Guide*. New York, 1995.
Newkirk, D. *Almanac of Soviet Manned Space Flight*. Houston, TX, 1990.

Space Exploration, Moon, Manned

The idea of lunar exploration goes back at least as far as 1638, when the protagonist of Francis Godwin's *Man in the Moon* rode a flock of 25 geese from the earth to the lunar surface. Animals played a rather less fanciful role in man's actual first steps into space, with the dog Laïka riding aloft in the Soviet spacecraft Sputnik-2 in November 1957. Early Soviet space successes prompted the U.S. to accelerate its own space program. The U.S. entered space with the launch of the unmanned probe Explorer-1 on 31 January 1958, and the National Aeronautics and Space Administration (NASA), was founded in October 1958. Manned moon exploration was preceded by a series of unmanned missions by both the USSR and the U.S. The USSR's Luna series of spacecraft successfully reached the moon as early as 1959; the first spacecraft to land on the moon was Luna-2, which touched down on 14 September 1959. The Soviet Luna-9 landed on the moon on 3 February 1966, transmitting pictures of the moon's surface back to earth. The U.S. followed suit with Surveyor-1, which achieved a soft lunar landing on the plain Oceanus Procellarum on 2 June 1966.

Manned space exploration began with the flight of the Russian Yuri Gagarin on 12 April 1961. It continued with the U.S. Mercury program, one of three successive U.S. programs leading up to the

first moon landing. The one-manned Mercury capsule was launched into space aboard a Mercury Redstone (MR3) launch vehicle. The first Mercury flight was accomplished by astronaut Alan B. Shepard, Jr. on 5 May 1961. The Gemini program followed. A major goal of the Gemini program was to increase the length of manned space flights. The two-man Gemini craft was lifted into orbit aboard a Titan-2 launch vehicle. Milestones in the Gemini program included the first rendezvous of one spacecraft with another, the first docking maneuver of two spacecraft, and the first "spacewalk" by American astronaut Ed White in June 1965 as part of the Gemini-4 mission (Figure 18). Information gathered on Gemini missions was critical in the Apollo program that would land the first man on the moon in July of 1969. The Apollo missions utilized a three-man space capsule carried into space aboard either a Saturn-4 or Saturn-5 booster rocket. The Apollo spacecraft had three components: the command module (CM) designed to return the astronauts to earth, the service module, housing the electrical and propulsion systems, and the insect-like lunar excursion module (LEM) designed to carry astronauts to the surface of the moon. The 4.6-meter-tall LEM weighed 12 tons, the 7-meter-tall service module, 25 tons. The crew comprised a commander, a LEM pilot, and a CM pilot. The CM pilot remained in orbit around the moon, while the commander and LEM pilot descended to the surface.

There were ten manned flights from 1968 to 1972. The first of the planned Apollo missions ended in tragedy on 27 January 27 1967, when a fire during a launch-pad test killed astronauts Roger Chaffee, Virgil Grissom, and Edward White. The disaster prompted a thorough re-design of the lunar and command modules. The first manned spacecraft to leave the earth's orbit and orbit the moon was Apollo-8, which began its lunar orbit on 24 December 1968. Apollo-10, launched on 18 May 1969, was the "dress rehearsal" for the actual moon landing that followed with Apollo-11. On this mission, which came within 14.5 kilometers of the moon, astronauts investigated the Sea of Tranquility and provided the first live color television broadcast from space. Two months later, Apollo-11 carried the first astronaut to the surface of the moon. Launched on 16 July 1969, the spacecraft landed on Sunday, 20 July at 4:17 p.m. Eastern Daylight Time in the Sea of Tranquility. Aboard the craft were Neil A. Armstrong, the first man to set foot on the moon, Edwin E. "Buzz" Aldrin, Jr., and Michael Collins. At 10:10 p.m., after sleeping for a few

Figure 18. Ed White, first American spacewalker, June 3, 1965.
[*Source: Courtesy of NASA.*]

Figure 19. Astronaut Eugene A. Cernan, Apollo-17 mission commander, makes a short checkout of the lunar roving vehicle during the early part of the first Apollo-17 extravehicular activity. This view of the "stripped down" Rover is prior to load-up. The mountain in the right background is the east end of South Massif.
[*Credit: Courtesy of NASA. Photograph by Geologist-Astronaut Harrison H. Schmitt.*]

hours, Armstrong and Aldrin walked on the moon as Collins remained aboard the command module, Columbia. Armstrong and Aldrin raised the American flag for the television camera and left it behind along with a few instruments and a plaque reading "Here men from planet earth first set foot upon the moon. July 1969 AD. We came in peace for all mankind." It is estimated that roughly 600 million people viewed the televised landing. The lunar module left the moon on 21 July after more than 21 hours on its surface, and the spacecraft returned to earth on 24 July with drilled core samples, photographs and moon rocks.

The Apollo program continued after the July 1969 landing, with the final mission, Apollo-17, launched on 7 December 1972 (see Figure 19). Like the Apollo-11 mission, Apollo-17 astronauts Eugene A. Cernan, Ronald E. Evans, and Harrison H. "Jack" Schmitt left behind a plaque, which read "Here man completed his first exploration of the moon, December 1972 AD. May the spirit of peace in which we came be reflected in the lives of all mankind." In total, the Apollo program executed three earth-orbiting missions, two lunar-orbiting missions, one lunar flyby, and six lunar landings.

See also **Space Exploration, Manned Orbiters; Space Exploration, Unmanned; Space Exploration: Moon, Unmanned Space Exploration, Planetary Landers; Space Launch Vehicles**

TIMOTHY S. BROWN

Further Reading

Chaiken, A. *A Man on the Moon*. Viking Press, New York, 1994.

Collins, M. *Liftoff*. Grove Press, New York, 1988.

Emme, E.M. *A History of Space Flight*. Holt, Rinehart & Wilson, New York, 1965.

Kennan, E.A. and Harvey Jr., E.H. *Mission to the Moon*. William Morrow & Co, New York, 1969.

Murray, C. and Cox, C.B. *Apollo: The Race to the Moon*. Simon & Schuster, New York, 1989.

Space Exploration, Moon, Unmanned

The exploration of the moon by unmanned spacecraft was pursued in phases (flyby, impact, soft landing and orbital injection), which represented increasing accuracy in guidance and control (see Table 1). The first spacecraft to fly near the moon, which is about 400,000 kilometers from earth, was the Soviet Luna-1, launched in January 1959. It passed within 6000 kilometers of the lunar surface and, in doing so, became the first man-made object to escape the earth's gravitational field. In September 1959, Luna-2 became the first spacecraft

Table 1 Lunar spacecraft targeting accuracy: progression from flyby and impact to orbit and landing (selected missions).

Launch	Spacecraft (state)	Achievement
Oct. 1958	Pioneer-1 (U.S.)	Reached distance of 113,854 km (less than one-third of the distance to the moon)
Jan. 1959	Luna-1 (USSR)	Passed within 5,955 km; escaped earth's field
Sept. 1959	Luna-2 (USSR)	First to impact moon
Oct. 1959	Luna-3 (USSR)	First far-side photos/TV images
Apr. 1962	Ranger-4(U.S.)	Far-side impact/spacecraft inoperative
Jul. 1964	Ranger-7 (U.S.)	4,308 photos/impact (first Ranger success)
Jan. 1966	Luna-9 (USSR)	First soft-landing; transmitted photos for 3 days
Apr. 1966	Luna-10 (USSR)	First to enter lunar orbit
May 1966	Surveyor-1 (U.S.)	Soft-landed 2 June 1966, 11,150 photographs
Aug. 1966	Lunar Orbiter-1 (U.S.)	First US orbiter/lunar mapper
Nov. 1970	Luna-17 (USSR)	First automated rover, Lunokhod-1

to hit the moon (note that only successful Soviet probes were given numbers). The mission's main intention—in an intensely competitive period known as the "space race"—was to deliver a 26-kilogram sphere carrying tiny hammer-and-sickle medallions.

The arguably more useful Luna-3, launched in October 1959, was placed in a large elliptical earth orbit which caused it to loop around the moon every 16.2 days. This allowed the Soviet Union another "first" in that it was able to photograph the lunar far side for 40 minutes on its first orbit. The pictures showed that there were many more craters and fewer maria, the large, flat, dark areas on the moon's surface, than on the near side. This was a significant technical achievement since the lunar far side is not visible from earth.

The technology involved in producing those early space photographs makes modern-day electronic systems look simple. Luna-3 was, in effect, a space-based photoprocessing laboratory required to operate under the weightless conditions and wide temperature extremes for which space is renowned. After exposure, the film was automatically developed, fixed and dried, then placed in front of a television scanning system, which converted the photographs into a stream of telemetry data. Once received on earth, this was converted back to a hard-copy photograph.

Following a number of failures, America too succeeded in lunar impact with one of its Ranger spacecraft; it was, however, inoperative at the time. Early Rangers carried a package of science experiments including a cosmic dust detector, magnetometer and x-ray scintillation counters. Later spacecraft (see Figure 20) were redesigned to support the Apollo manned lunar program by returning close-up photographs of the surface. The Ranger program celebrated its first success in July 1964, when Ranger-7 returned 4308 photographs of the lunar surface before impact.

It was the Soviet Union, however, that would perform the first soft landing with Luna-9 in January 1966. Luna-9 was a fairly sophisticated spacecraft comprising two elements, a mother-craft and a lander. The lander, a sphere with four petal-shaped panels which opened after landing to right the spacecraft and point the integral antennas towards earth, carried an instrument package which included a television camera and an ingenious deployable rod with mirrored surfaces which acted as a "rear-view mirror." The spacecraft sent back panoramic views of its surroundings, and measured a temperature of 300°C and a

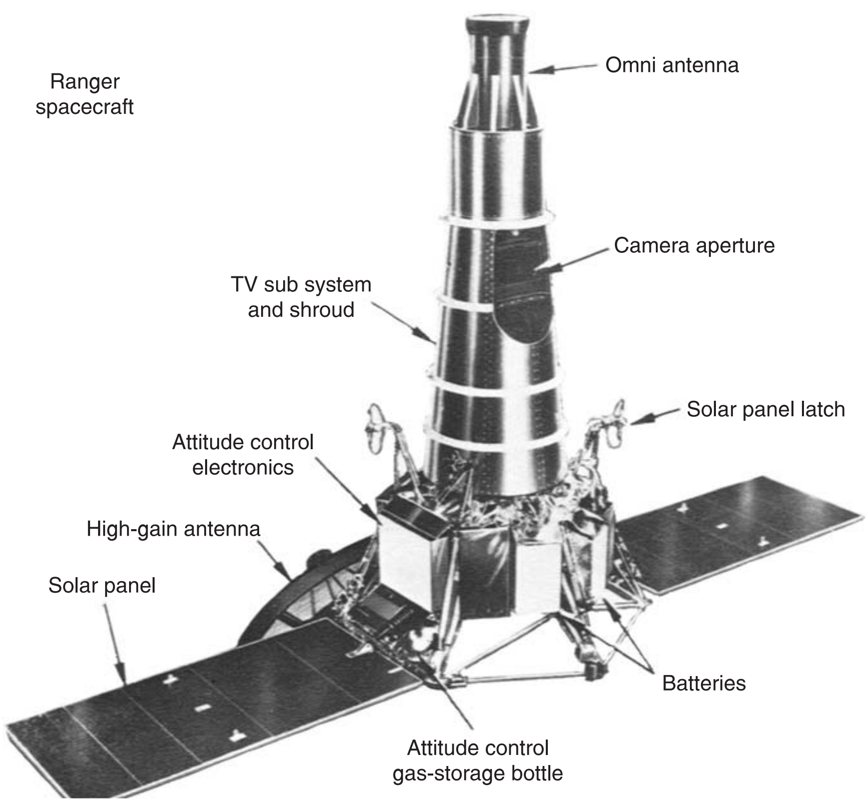


Figure 20. NASA's Ranger spacecraft showing principal equipment. The solar panels were folded against the sides of the spacecraft for launch and deployed once en route to the moon; the circular high-gain antenna (part hidden) was folded up against the base of the spacecraft. [*Source: Courtesy of NASA*.]

radiation level of 30 millirads per day. Importantly, it proved once and for all that a lunar lander would not sink beneath a sea of dust, as some had hypothesized.

After Luna-10 became the first spacecraft to enter lunar orbit in April 1966, the four phases of unmanned lunar exploration were complete. Between August 1966 and August 1967, America succeeded in placing five spacecraft in lunar orbit. The aptly-named Lunar Orbiters returned a wealth of high-quality imagery from orbits between 1600 kilometers and 40 kilometers above the surface. As with earlier spacecraft, the Lunar Orbiters took pictures on film, developed them automatically, then scanned them into a TV system for transmission to earth. The system produced images with some 500,000 times the resolution of the Ranger cameras and showed features measuring about 3 meters across. The fact that the best resolution of earth-based telescopes of the time was about 800 meters proved the advantage of sending spacecraft there. The Lunar Orbiters also discovered mysterious large concentrations of mass—dubbed "mascons"—below the lunar maria. These proved to be extremely important for predicting the orbital paths of the later Apollo spacecraft.

NASA's Surveyor program, operated in parallel with the Orbiter missions, succeeded in soft-landing five out of seven spacecraft, providing a reasonable level of confidence that an Apollo lunar module could be guided to a safe and accurate landing. In the context of the history of technology, it is noteworthy that by mid-1966, when slide-rules and rotary-dial telephones were still the norm, both Russia and America had succeeded in soft-landing spacecraft on the moon.

In November 1970, following the Apollo-11 and -12 manned lunar landings (see Figure 21), the USSR succeeded in delivering the first of two teleoperated rovers, Lunokhod-1, to the moon; Luna-21 deployed Lunokhod-2 in January 1973.

Considering the importance attached to lunar exploration in the 1960s, it is significant that no lunar missions were conducted between Luna-24's landing in August 1976 and the launch of Japan's Hiten/Hagoromo spacecraft in January 1990. This was followed by two U.S. lunar mapping missions—Clementine in 1994 and Lunar Prospector in 1998—the most significant result of which was the suspected detection of water ice at the lunar poles. Although there were plans at the turn of the century to launch further unmanned missions to the moon, it seemed unlikely that any serious lunar exploration would ensue until a decision is made to mount a further manned program of exploration and development.

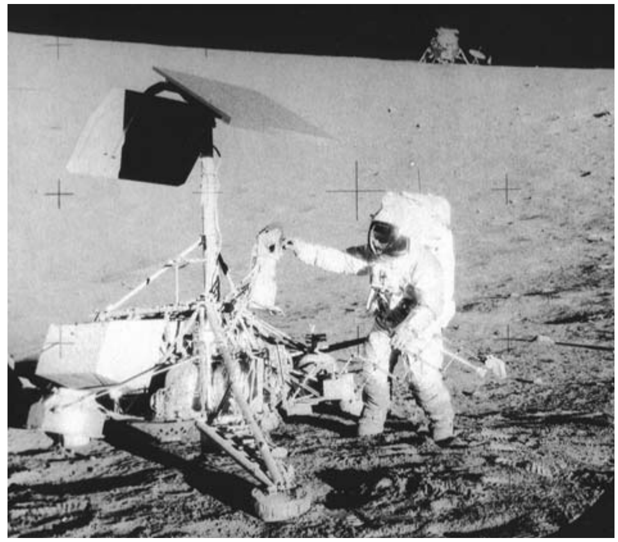


Figure 21. Apollo-12 astronaut Pete Conrad inspects Surveyor-3 two-and-a-half years after it soft-landed on the moon. Landing demonstrations with the Surveyors helped to refine the techniques that would allow Apollo-12 to land within 200 meters of the spacecraft (the lunar module Intrepid is visible on the horizon). Parts of the Surveyor were returned to earth for analysis.
[*Source: Courtesy of NASA*.]

See also **Space Exploration, Moon, Manned; Space Exploration, Unmanned**

MARK WILLIAMSON

Further Reading

Caidin, M. *Race for the Moon*. William Kimber, London, 1960.
Gatland, K. *Robot Explorers*. Blandford Press, London, 1972.
Kopal, Z. *Exploration of the Moon by Spacecraft*. Oliver & Boyd, Edinburgh, 1968.
Reeves, R. *The Superpower Space Race*. Plenum Press, New York, 1994.
Spudis, P.D. *The Once and Future Moon*. Smithsonian Institution Press, Washington, 1996.
Troebst, C.-C. *Reaching for the Moon*. Hodder & Stoughton, London, 1961.
Wilson, A. *Solar System Log*. Jane's, London, 1987.
Yenne, B. *The Encyclopedia of US Spacecraft*. Hamlyn, Twickenham, 1985.

Space Exploration, Planetary Landers

With the launch of the first earth-orbiting satellites, the use of unmanned planetary landers, probes, and telescopes came under consideration as a means to chart and explore the solar system. The

advantages of such machines became clear immediately, as they represented a short cut in the race between the U.S. and the Soviet Union to dominate this new realm of international competition. The design and shape of these instruments varied but displayed similar basic requirements: the maximum use of the area within the superstructure for the instruments and the onboard computer, a means to communicate with earth to receive guiding instructions and send back information collected, camera and sensor pods, and an engine and its fuel capable of functioning for several years.

The first "target" was earth's moon, and both superpowers launched probes beginning in 1958. The first "wave" did not involve landers as such, but machines intended to orbit and perhaps crash on the moon. None of the American Pioneer machines succeeded in this task, but two Soviet Luna spacecrafts came close to earth's satellite. The U.S. then initiated the Ranger project. Begun in 1959, the first four missions failed outright, but Ranger-7 succeeded in sending back photographs of the lunar surface, which paved the way for the Apollo manned moon landings.

The exploration of Venus also began with the Cold War as background. The peculiarity of Venus' cloudy atmosphere as well as its impact on human culture dating back to the earliest civilization made it an important target. The Soviet Union first launched probe Venera-1 on 12 February 1961, but the probe failed on its way to Venus. The U.S. followed suit in summer 1962 with Mariner-1 and Mariner-2. The first failed, but the second one was able to take atmospheric measurements. Several other Venera and Mariner missions followed, all confirming that the planet was completely inhospitable and that the greenhouse effect actually had elevated ambient surface temperature to 430°C. The most recent planetary probe was Magellan, launched in 1989, which mapped most of the planet at high resolution.

Mars drew an equal or even greater interest in its exploration. The Soviet Union reached Mars first, but failed to get any data back from its June 1963 mission. The American Mariner-4 mission flew near the planet in summer 1965, and Mariner-6 and Mariner-7 further mapped the Mars surface in 1969 in preparation for a landing. None of the probes suggested there was any life on Mars. Exploration nevertheless reached a zenith with the Viking-1 and Viking-2 missions. Launched 10 weeks apart in 1975, both machines took about one year to reach Mars. Once there, the landers' missions confirmed the lack of life on Mars due to the high level of ultraviolet radiation that saturates the Martian atmosphere. Although the Viking program ended in November 1976, the probes continued to transmit for another six years. Further planetary probes explored Mars and its satellites. The most successful was the Mars Pathfinder which landed on July 4, 1997, and whose robotic rover, Sojourner, spent 30 days collecting data on the environment and relief of the planet.

The distance between earth and other planets in the solar system has prompted the need for other types of exploration, including solar system probes and space telescopes. The first attempt at surveying outer planets involved the preparation of the Pioneer-10 and Pioneer-11 probes. The first, launched in March 1972, reached the vicinity of Jupiter 21 months later, then pursued a course out of the solar system. Affixed to its side was a plate with a male and female figure and a series of symbols depicting Voyager's origins and travel path. The plaque represented an attempt at considering the possibilities in the search for extraterrestrial intelligence (SETI). Although it was elegantly simple, the design was criticized for depicting the female figure in a passive stance whereas the male raises his hand in greeting, and both figures appear to be Caucasian. As for Pioneer-11, it passed near Saturn twice, and in 1990 left the solar system.

Two other American probes, Voyager-1 and Voyager-2, launched in 1977, conducted a planetary tour of the two outermost planets, Uranus and Neptune, and 48 moons. The last investigation of Jupiter was that of the probe Galileo. Although slower than other probes (it used the gravity of planetary bodies to correct its course and increase its speed, also known as the "slingshot" effect), Galileo reached Jupiter in 1995, six years after its launch. It successfully parachuted a probe into the atmosphere that transmitted for 45 minutes before the atmospheric pressure and the planet's gravity broke it up.

An alternative to expensive probes and the problems of earth-based telescopes (which include visual distortions through the atmosphere) has involved satellite telescopes. Several National Aeronautics and Space Administration (NASA) projects were devoted to this alternative beginning with the Apollo moon missions. The advanced orbiting solar observatory (AOSO), initially scheduled for Apollo moon missions, was turned into the Apollo telescope mount (ATM) and launched aboard the American Skylab space station in 1973. It functioned throughout the station's manned

operations (a total of some six months) and recorded solar activity.

Meanwhile, research into a giant space telescope had begun in the 1960s. The NASA project, which later became known as the Hubble space telescope (HST), capitalized on existing technical knowledge used in the manufacture of spy satellites to order a high-resolution mirror. Although the system underwent several redesigns to fit into a space shuttle bay (instead of the top of a Saturn-V moon rocket as had been initially planned), it was ready for launch in the mid-1980s. The shuttle Challenger accident of January 1986 delayed the launch date by four years. Once lofted, however, the HST displayed an aberration in its picture resolution. The error, caused by the improper position of one of the mirrors, required a "corrective lens" that was installed during a 1993 shuttle-servicing mission. The repair was successful and helped HST uncover new information about black holes.

See also **Space Exploration, Manned Orbiters; Space Exploration, Moon, Manned; Space Exploration, Unmanned; Space Stations, Skylab**

GUILLAUME DE SYON

Further Reading

Ezell, E. *On Mars: The Exploration of the Red Planet, 1958–1978*. NASA, Washington D. C., 1984.

Fimmel, R.O., Swindell, W. and Burgess, E. *Pioneer Odyssey*, NASA, Washington D.C., 1977.

Hall, C. Lunar Impact: A History of Project Ranger. NASA, Washington D.C., 1977.

Nicks, O.W. *Far Travelers. The Exploring Machines*. NASA, Washington D.C., 1985.

Smith, R.W. *The Space Telescope: A Study of NASA, Science, Technology, and Politics*. Cambridge University Press, New York, 1994.

Space Exploration, Unmanned

Manned space exploration receives most of the glory and the attention, but unpiloted spacecraft have been used as tools longer and for a greater variety of reasons. Some even believe that humankind should cease piloted space exploration in favor of the much less expensive unpiloted probes when we consider the accomplishments of these robotic craft.

The beginnings of unpiloted space exploration were made by the Soviets. The first artificial earth satellite, Sputnik, or "Fellow Traveler" carried only a radio transmitter and batteries and no other instruments despite its 80-kilogram-plus orbiting weight. Its batteries lasted about two weeks after the 4 October 1957 launch. The Soviets had been working on an earth-orbiting laboratory, Object D, as their first satellite, but Sergey Korolev, the most influential chief designer in their program, was anxious that the U.S. would launch one first. He thus convinced his colleagues to prepare a couple of "simple satellites" for orbit, which eventually became Sputniks-I and -II. Object D was thus pushed down to third in the launch list. When the U.S. orbited its Explorer-I satellite on 31 January 1958, it carried a radiation detector and some other small instruments. The radiation detector picked up evidence of a radiation belt around the earth. Object D, when it became Sputnik-3, confirmed this. The era of robotic exploration had begun.

Several dozen Explorers and hundreds of Soviet Cosmos satellites have explored near-earth space, but bodies such as the moon, Mars, and Venus have attracted the most attention from the public. The moon came under early and frequent exploration. After three American failures in 1958, the Soviet Luna-1 flew past in 1959, discovering the solar wind, and the American Pioneer-4 was launched. The Soviets followed up that year with a hard lander, Luna-2, the first person-built object to land on another celestial object, and Luna-3, which imaged the lunar far side. With Ranger-4, the Americans hit the moon. By then, President John F. Kennedy had pledged America to land a person on the moon and return them to earth by 1970, so the unpiloted programs took on the air of being preparatory to the main event. The Rangers, which flew until 1965, impacted the moon taking pictures all the way down. The Soviets had by then turned their interest to soft landers, suffering several failures that made them into lunar impactors. The Soviets successfully launched a Lander and orbiter to the moon early in 1966. The American Surveyor series began landings soon after. In the summer of 1966 the first American Orbiter arrived at the moon.

During the next two years, America, convinced that it was in a race to make the first piloted lunar landing, sent seven Surveyors, and five Orbiters to the moon, all but a couple of Surveyors successfully. The Soviets continued to fly orbiters, then graduated to lunar fly-arounds with returns to earth on two Zond missions in the second half of 1968, seemingly in further rehearsal of piloted mission. However, the Americans orbited the moon with a piloted mission near Christmas and landed on 20 July 1969. Any race, if there was one, was over.

The Soviets then turned to sample return missions, in which soil samples from the lunar

surface would be returned to earth. The first successful one of these was Luna-16, which landed on 20 September 1970. This was followed over the next six years with several more missions, two with Lunokhod rovers. The similarly between the Soviet science accomplishments on their unpiloted missions and the accomplishments of American piloted missions lent support for those who opposed piloted space exploration as too expensive and dangerous for its worth.

After a 14-year hiatus, the Japanese Muses-A orbiters visited the moon the first month of 1990. This was the first non-U.S. or non-Soviet probe to reach the moon. The Japanese are expected to launch penetration missions early in the 21st century. Meanwhile, the U.S. mounted two missions in the 1990s. The first was the Clementine, intended for technology demonstration and launched by the Ballistic Missile Defense Agency. The second was Lunar Prospector, one of the NASA's ''faster, cheaper, better'' missions searching for water and mineral deposits useful for eventual lunar bases occupied by humans. (''Faster, cheaper, better'' is an effort to avoid the excessive costs of space missions.)

Venus was the target of 17 Soviet spacecraft in the 22 years between 1961 and 1983 and five American overall. The early probes revealed Venus to be as inhospitable as it was beautiful. Mariner-2 discovered a 425°C surface temperature and a thick atmosphere. The Soviets decided to land on the planet, regardless. Their Venera-3 spacecraft was intended to explore the atmosphere, but communications failed just after atmospheric entry and it impacted, although this was quite good navigation for 1966. Then followed a series of spacecraft that returned atmospheric data before being crushed by the thick carbon dioxide atmosphere. First Venera-4 on 18 October 1967 returned data on the composition of the atmosphere and the surface temperature, now measured at 500°C. Venera-5 lasted until it reached 26 kilometers of the surface and its sister, Venera-6, went within 11 kilometers. Venera-7 made the first landing on another planet on 15 December 1970 and lasted 23 minutes.

The Soviets then made probes that landed and lasted longer and longer. They carried increasingly complex sensor suites, eventually including imaging equipment. Venera-8 measured wind speeds while descending and lasted 50 minutes after landing. Venera-9 and -10 were orbiters and landers. Venera-9 worked on the surface for 53 minutes, including sending the first picture from another world on 25 November 1975. Venera-10's lander achieved 65 minutes of life on the surface.

The Americans, after a couple of Mariner fly-past missions in the 1960s, sent a pair of Pioneers in 1978. The first, Pioneer-12, was an orbiter with a radar-mapping device that lasted until 1992. The second, Pioneer-13, released a cluster of four atmospheric probes. The Soviets also sent two probes during the 1978 launch window (the period when the planets are best aligned), reaching 95 and 110 minutes of surface life.

The Soviets ended their exploration of Venus with four straight successful probes, Venera-13, -14, -15, and -16. The first pair were fly-pasts with landers, the last two orbiters and landers, all had color imaging equipment, and the orbiter returned a map of Venus's northern hemisphere. Mapping was the chief goal for the 1989–1994 mission by the Magellan American spacecraft that used its synthetic aperture radar on 99 percent of the planet's surface. It was the final probe to reach Venus.

In contrast to the relative success exploring the inhospitable Venus, probing relatively benign Mars proved problematic. Only three of sixteen Soviet probes of Mars have seen full success, while eight of fourteen American probes were satisfactory; less than a third of the combined missions were thus successful. After five Soviet failures and the failure of its partner, Mariner-III, Mariner-IV flew past the Red Planet and returned pictures of craters on Mars. After eleven years and eight failed missions, the Soviet Mars-3 soft-landed on 2 December 1971. Even still, it relayed only 20 seconds of data to its orbiter, which lasted nine months. After a few more fly-pasts, the US Mariner-IX went into Martian orbit at about the same time as Mars-3. A planet-wide dust storm was in progress, and the temperature variations from it inspired the concept of a nuclear winter that scientists theorized would follow any widespread atomic exchange. The Soviet Union succeeded with another orbiter, Mars-5, after the failure of Mars-4, to survey possible Mars-6 and -7 landing sites. However, both probes failed.

The U.S. then had a pair of highly successful orbiters and landers, Viking-1 and -2, which arrived during the 1976 launch window and relayed data until 1980. After over a decade of hiatus, the Soviets tried again with Phobos-1 and -2 in 1988, both including a Phobos lander. Both failed, as did the next American mission, Mars Observer. A replacement spacecraft, Mars Global Surveyor, was launched in November 1996, and entered Martian orbit the next year. The Russians attempted the ambitious Mars-96 mission that same launch window, an orbiter, two landers, and two soil penetrators. The booster failed,

leaving the spacecraft to crash into the ocean with 270 grams of plutonium intended to generate electrical power.

The American Mars Pathfinder, one of NASA's "faster, cheaper, better" missions, landed on Mars and released a rover, *Sojourner*. This mission returned a torrent of data on the Martian surface for 83 days after its arrival 4 July 1997. America then suffered a series of embarrassing failures at Mars. The first was Mars Climate Orbiter, which burned up on approaching the planet due to one of its sub-teams calculating in metric units and the other in imperial units. This was followed by the loss of the Mars Polar Lander and its Deep Space-2 Penetrators. All of these Soviet and U.S. failures prompted a panel at a NASA conference, entitled "Why is Mars So Hard?"

Japan sent its Nozomi probe on its way to Mars in 1998, but at the time of writing, it had not arrived. Whether successful or not, it is, however, the first interplanetary probe launched by a country other than Russia or the U.S.

Looking at the deepest end of space exploration is NASA's Great Observatories program, which includes infrared and x-ray observatories, but the most famous is the optical Hubble space telescope. It was placed into orbit by the Space Shuttle on 25 April 1990, and serviced several times, each visit yielding a better spacecraft. It has mapped the entire sky and seen matter at the origins of the universe. It places the exploration of the Moon, Venus, and Mars essentially in our back yard.

See also **Space Exploration, Fly Past; Space Exploration: Moon, Unmanned; Space Stations, International Space Stations; Space Stations, Mir; Space Stations, Skylab; Telescopes, Space**

JAMES E. TOMAYKO

Further Reading

Anderson, Jr., F.W. *Orders of Magnitude: A History of NACA and NASA*, 1915–1980, NASA, Washington, D.C., 1981.

Ezell, E. *On Mars: The Exploration of the Red Planet, 1958–1978*. NASA, Washington D.C., 1984.

Sheldon II, C. *Review of the Soviet Space Program*. McGraw-Hill, New York, 1968.

Washburn, M. *Distant Encounters: The Exploration of Jupiter and Saturn*. Harcourt Brace Jovanovich, New York, 1984.

Useful Websites

Hamilton, C. *History of Space Exploration*: http://www.solarviews.com/eng/history.htm

NASA Headquarters homepage: www.hq.nasa.gov

Space Launch Vehicles

A launch vehicle is a rocket-based vehicle designed to carry payloads into space. Theoretically, there are two main types: expendable launch vehicles (ELVs) comprising a number of propulsive stages which are used and then discarded; and reusable launch vehicles which return from space intact to be refueled for another mission. Although no fully reusable launch vehicles were developed in the twentieth century, the Space Shuttle was designed to be partially reusable in that the orbiter returns to earth and the solid rocket boosters can be refurbished and refueled for subsequent flights.

Following Goddard's 1926 launch of the world's first liquid-fuel rocket, rockets of all types were developed for both military and civilian purposes. As with other technologies, World War II produced a dramatic leap in rocket technology, culminating in the V-2 ballistic missile. After the war, the V-2 and its designers became the basis of both American and Soviet missile research, leading eventually to the intercontinental ballistic missile (ICBM). However, the ultimate spur to the development of space launch vehicles came from the USSR, when on 4 October 1957 it launched Sputnik-1 into orbit. This fuelled the "space race" of the 1960s and the conversion of ICBMs into launch vehicles for manned and unmanned spacecraft.

Apart from improvements in reliability, accuracy and control, the main aim of launch vehicle development since the beginning of the space age has been to increase the payload mass that can be delivered to space. The payload capability of the early vehicles was low, especially the American rockets which were much smaller and less powerful than their Soviet counterparts. This is illustrated by the fact that America's first satellite, Explorer-1, weighed just 8.3 kilograms compared with Sputnik-1's mass of 83.6 kilograms; and whereas America's second satellite, Vanguard, launched by an even smaller rocket, weighed 1.5 kilograms, the USSR's Sputnik-2—carrying the dog Laïka—weighed over 508 kilograms. Little wonder that Soviet Premier Khrushchev nicknamed Vanguard "the grapefruit."

Progress was rapid in the 1960s, however, and by 1968 America had succeeded in sending the Apollo-8 spacecraft weighing more than 30,000 kilograms around the Moon. This was due entirely to the development of the Saturn-V, the first of what came to be known as heavy-lift launch vehicles. Modern launch vehicles can be divided into four groups based on payload capability: small vehicles which can deliver up to about 1000

kilograms to geostationary transfer orbit; medium vehicles (1000 to 2000 kilograms); large vehicles (2000 to 5000 kilograms); and heavy-lift vehicles (more than 5000 kilograms).

The launch vehicle is fundamental to space exploration and development: it is the active element of a space transportation system, in the same way that an aircraft is to an airline. This was made especially clear when the field evolved from government-operated launch vehicles launching scientific and military satellites to commercially operated vehicles launching fleets of commercial communications satellites.

In the U.S., former government-operated ELVs such as the ICBM-based Atlas, Delta, and Titan were commercialized when the Space Shuttle failed to live up to its promise. First launched in April 1981, the Shuttle had been expected to provide cheaper and more reliable access to space, effectively replacing the old-style ELVs. However, following the Challenger accident in January 1986 it was decided not to use it for commercial payloads and the ELVs were given a new lease of life. In fact, the American ELVs in use at the close of the century were essentially upgrades of 1950s ICBM technology.

The European Space Agency (ESA), by contrast, decided to develop an ELV specifically as a commercial satellite launcher; it was named Ariane and conducted its first launch in December 1979. By the end of the century, its fifth variant, Ariane-5 (see Figure 22), was in operation, along with a number of other commercial, or semicommercial, launch vehicles from China, Japan, India and—following the end of the Cold War—from Russia and Ukraine. Several of these vehicles now compete for each satellite launch contract in an international market analogous to the air-cargo business.

Figure 22. A cutaway of the Ariane-5 launch vehicle showing two satellites under the payload fairing and the strap-on solid propellant motors.
[*Reproduced with permission from Arianespace*.]

Most launch vehicles use liquid propellants, because they generally provide better performance than solid propellants. However, solids are often utilized in strap-on boosters, which augment the thrust of a given launch vehicle variant and improve its payload capability. Most launch vehicles employed to deliver payloads to orbit (as opposed to suborbital trajectories) have more than one stage, each with its own engines and propellant tanks. The most common design has three stages, although both two- and four-stage rockets exist.

The main advantage of the multistage rocket (see Figure 23) is that empty tanks and associated structure do not have to be carried all the way to space, with the result that a larger payload can be placed in a given orbit. The disadvantage of the staged design is that most, if not all, of the vehicle is discarded on every mission, which is rather like scrapping an airliner after a single flight.

It has not escaped the notice of the space industry that this is an extremely wasteful and expensive method of accessing space and there have been many proposals, detailed designs and even technology demonstrations aimed at the development of a fully reusable launch vehicle. However, severe technical difficulties associated with the development of appropriate propulsion systems, materials and aerodynamic designs, and most of all the high cost of these developments, have so far obstructed all attempts to find a solution. Vertically, space may only be 100 kilometers away, but it has proved to be one of the

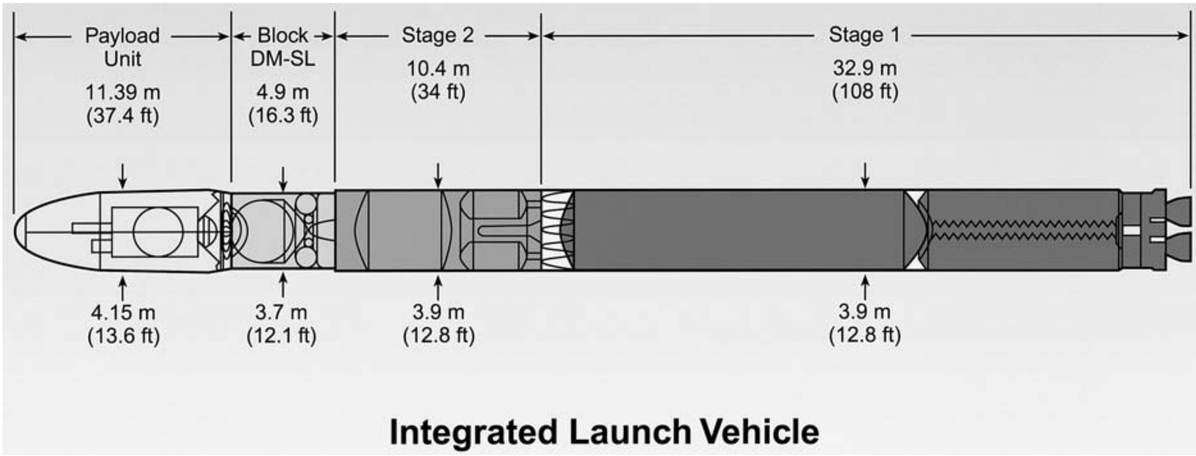


Figure 23. The stages of the Sea Launch expendable launch vehicle showing (from left to right) the payload accommodation, the Block DM-SL upper stage, the second stage and the first stage.
[*Reproduced with permission from Boeing*.]

most challenging journeys technology has ever been required to make.

See also **Rocket Planes; Rocket Propulsion, Liquid Propellant; Rocket Propulsion, Solid Propellant**

MARK WILLIAMSON

Further Reading

Baker, D. *The Rocket: the History and Development of Rocket and Missile Technology*. Crown Publishers, New York, 1978.

Chiulli, R.M., Ed. *International Launch Site Guide*. Aerospace Corporation, El Segundo, 1994.

Hill, C. N. A *Vertical Empire: the History of the UK Rocket and Space Programme, 1950–1971*. Imperial College Press, London, 2001.

Isakowitz, S.J. *International Reference Guide to Space Launch Systems*, American Institute of Aeronautics and Astronautics, Washington D.C., 1995.

Ordway, F.I. and Sharpe, M.R. *The Rocket Team*. Heinemann, London, 1979.

Rycroft, M., Ed. *The Space Transportation Market: Evolution or Revolution*? Kluwer, Dordrecht, 2000.

Stoiko, M. *Soviet Rocketry: The First Decade of Achievement*. David & Charles, Newton Abbot, 1970.

Space Shuttle

The Space Shuttle was developed by the National Aeronautics and Space Administration (NASA) as one component of a program intended to secure the future of the agency after the lunar landings in 1969 and to ensure that the U.S. maintained world leadership and technological superiority in space. Initially conceived as a system to transport humans and materials to a space station, its final configuration embodied a trade-off between NASA's ambitions and the economic and political constraints ensuing on a loss of support for space among the American people and Congress in the early 1970s.

The Shuttle system was intended to be fully reusable, as opposed to the more conventional, expendable launch vehicles (ELVs) or rockets, and was supposed to replace them. In the end it was only partly reusable. NASA had to accept design features demanded by the Department of Defense (DOD) to secure the military's support for the system, and within a few years of its becoming operational the DOD were once again demanding that America's access to space be secured with a mixed fleet of vehicles, including ELVs, striking a further blow to the Shuttle's commercial viability.

Four main components comprised the Shuttle system. First, there was the orbiter itself. Five have been built, Columbia, Challenger (replaced by Endeavor), Discovery and Atlantis (all named after famous exploration sailing ships). The reusable orbiter itself was akin to a delta-winged space plane, which could be brought back to earth and landed on a runway, like a conventional aircraft. It comprised a large crew compartment, a cargo bay about 4.5 meters in diameter and 18 meters long, and three main engines. It could transport up to 29,500 kilograms into near-earth orbit (185 to 1111 kilometers). The orbiter was multipurpose, and in addition to carrying personnel and payloads into space, it could serve as an orbiting service center for other satellites (as in the space walks to repair the defects in the Hubble space telescope) and to return previously orbited spacecraft to earth. Its cargo bay

has also housed a laboratory called Spacelab. This was built by the European Space Agency with major funding from Germany, and provided a shirt-sleeve environment for performing scientific experiments under conditions of microgravity.

Fuel for the orbiter's main engines was supplied from an external tank that also served as the structural backbone for the system during launch operations. It was 46.8 meters long and 8.4 meters in diameter. Its gross weight at liftoff was over 750,000 kilograms and it contained almost 543,000 liters of liquid oxygen and almost 1.5 million liters of liquid hydrogen. These propellants supplied the orbiter's three main engines for about eight minutes, whereupon they cut off, at an altitude of about 109 kilometers, just before the spacecraft was injected into low-earth orbit. The orbiter separated from the external tank, which followed a ballistic orbit, and splashed down into the ocean. This tank was not recovered.

Two solid-fuel boosters were strapped on to the sides of the external tank. Each was 45.5 meters long and 3.7 meters in diameter, and contained 500,000 kilograms of propellant (a mixture containing about 70 percent ammonium perchlorate). Together with the orbiters' three liquid-fuelled engines these boosters provided over 3 million kilograms of thrust to lift the Shuttle off the pad. They were designed to burn for just over 2 minutes, and at a height of about 44.5 kilometers they separated from the external tank, falling into the ocean at predetermined points where they were recovered for reuse.

When NASA promoted the Space Shuttle in the early 1970s, the Agency claimed that it would revolutionize space transport in the decade to come. Being reusable, costs would be slashed as compared to ELVs (whose use was compared to operating a railroad by throwing away the locomotive after each trip). Using mission models that predicted that the orbiter would be launched once a week, NASA and its consultants insisted that the Shuttle would make access to space "routine and economical." This proved to be wildly optimistic. Columbia, the first orbiter to be flown, took off from Cape Canaveral on 12 April 1981, landing successfully at the Edwards Air Force Base two days later. By 26 January 1986, when Challenger exploded just over a minute after lift-off, there had been only 24 launches in almost five years, about as many as NASA had originally claimed it could launch in six months. Cost per kilogram into orbit was also far higher than anticipated.

Several reasons account for the huge disparity between the initial estimates made by NASA and the Shuttle performance achieved in practice. The need to "sell" the system to a dollar-conscious Congress and Office of Management and Budget led the Agency to exaggerate the cost-effectiveness of its new space transportation system. There were also a number of uncertainties that are inherent in any new, revolutionary technological project and which only seem evident with hindsight. Above all, it took far longer to turn-around the reusable spacecraft because, being so complex, and being "man-rated," it took several months, rather than several days, to refurbish an orbiter that had withstood the rigors of space travel, and to be satisfied that it was flight-ready and safe for humans.

Even then, risks were always taken. Indeed the detailed analysis of the Challenger accident shows that flight engineers and managers necessarily had to make many micro-decisions about the level of acceptable risk in the process leading up to the launch. Those responsible for the orbiters' solid-fuel boosters knew that the O-rings between their segments could 'freeze' in the cold Florida morning, and not seal efficiently. They deemed this risk to be acceptable, in the light of their previous and ongoing studies of O-ring behavior. Their error cost the lives of seven astronauts, saddened millions of people in the U.S. and around the world, and struck a serious blow to NASA's reputation.

The Shuttle resumed operations in September 1988, and continues to be launched on an approximately bimonthly basis. The low launch frequency and cost, the Challenger accident, and NASA's decision to phase out ELVs created an opportunity for other launch vehicle suppliers to gain an important share of the market in the mid-1980s. Europe's Ariane rocket was a major benefactor, and today Arianespace has 50 percent of the world market. This is one of the ironies of history. The European program was embarked on in the shadow of NASA's insistence that ELVs would soon be obsolete, and it was only thanks to French political determination to break the U.S. monopoly on access to space, notwithstanding the risks, that this alternative to the Shuttle was developed.

See also **Space Exploration, Manned Orbiters; Space Stations, International Space Station; Space Stations, Mir; Space Stations, Skylab**

JOHN KRIGE

Further Reading

Heppenheimer, T.A. *The Space Shuttle Decision. NASA's Search for a Reusable Launch Vehicle*. NASA-SP4221, Washington D.C., 1999.

Krige, J. The decision taken in the early 1970s to develop an expendable European satellite launcher, in *A History of the European Space Agency, 1958–1987. Vol. II: The Story of ESA, 1973–1987*, Krige, J., Russo, A. and Sebesta, L., Eds. European Space Agency, Noordwijk, 2000.

Launius, R. *NASA: A History of the U.S. Civil Space Program*. Krieger, FL, 1994.

Vaughan, D. *The Challenger Launch Decision. Risky Technology, Culture and Deviance at NASA*. University of Chicago Press, Chicago, 1996.

Useful Websites

NASA Historical Data Books. Volume V. NASA Launch Systems, Space Transportation, Human Spaceflight, and Space Science, 1979–1988. NASA SP-4012, Washington D.C., 1999: http://history.nasa.gov/SP-4012/cover.html

Space Stations, International Space Station

The International Space Station (ISS) program is an international collaboration of the U.S., Europe, Japan, Canada, and Russia, each of which is expected to provide modules and equipment designed to support a crew of up to six for at least 15 years. The ISS will be the largest structure ever built in space: at completion its overall dimensions will be approximately 110 by 60 meters, about the size of a soccer pitch. Its mass will be some 420,000 kilograms and its total pressurized volume about 900 cubic meters, equivalent to the passenger cabins of two Boeing 747 aircraft.

U.S. President Reagan directed National Aeronautics Space Administration (NASA) to build a space station in 1984, but redesigns and cost overruns delayed the beginning of orbital assembly until November 1998. Over that period, its original name Freedom was changed first to Alpha and then to ISS, as the team of international partners was amassed. The first module to reach orbit, on 20 November 1998, was Zarya, formally known as the functional cargo block (or, from its Russian name, FGB). It is a self-contained spacecraft—a power, propulsion, and orbital-control module that is pressurized and thus provides habitable accommodation. A U.S.-built connection node called Unity was docked to Zarya on 7 December 1998. Then, following significant delays in its construction, the Russian-built service module Zvezda, a habitable command and control center, was added on 26 July 2000.

Modules expected to join the ISS in later years include the Columbus Orbital Facility (COF), a general-purpose science and technology laboratory supplied by the European Space Agency (ESA), and the Japanese Experiment Module (JEM) supplied by the Japanese space agency NASDA (see Figure 24).

Although the station is being assembled at an orbital altitude of about 380 kilometers, its operational orbit, attained late in the assembly sequence, is expected to have an altitude of 426 kilometers and an inclination of 51.6 degrees. The lower initial orbit allows the Space Shuttle to deliver some 18,000 kilograms of additional payload on each mission, while the inclination is designed to allow the station to over-fly some 85 percent of the earth's landmass and 95 percent of its population.

The six main modules and a number of connecting nodes will be mounted on an "integrated truss structure," the station's structural backbone, along with eight two-panel solar arrays designed to rotate to maximize their exposure to the sun and provide a total of some 110 kilowatts of power. The station's location in low-earth orbit means that it will be eclipsed by the planet on each orbit, obliging its systems to draw power from its nickel–hydrogen (NiH_2) batteries which will be recharged from the arrays during the sunlit portion of the orbit.

The station's orbital track and orientation, or attitude, was determined initially by gyroscopes alone but later used an inertial navigation system incorporating global positioning system (GPS) receivers. Attitude control is engineered using electrically powered momentum wheels, which exchange angular momentum with the station's structure to rotate it in any axis. In common with all unmanned spacecraft, such as communications satellites, adjustments can also be made by bipropellant thrusters, and backup systems are provided for all subsystems in case the primary equipment should fail. The station's altitude is boosted periodically, to overcome atmospheric drag, by firing rocket engines on visiting spacecraft.

In addition to the standard subsystems, there are several novel technological aspects to the station. One is the inclusion of a Canadian-built manipulator arm, which can move along the station's truss on a mobile transporter enabling the crew to perform assembly and maintenance work without leaving the safety of the modules. Although the arm is based on the Space Shuttle's remote manipulator system, it has a "hand" at either end and is not permanently fixed to the station's structure so that it can crawl along the truss. Also new for the ISS is a pair of cupola modules, supplied by ESA, each of which has seven relatively large windows to allow observations of

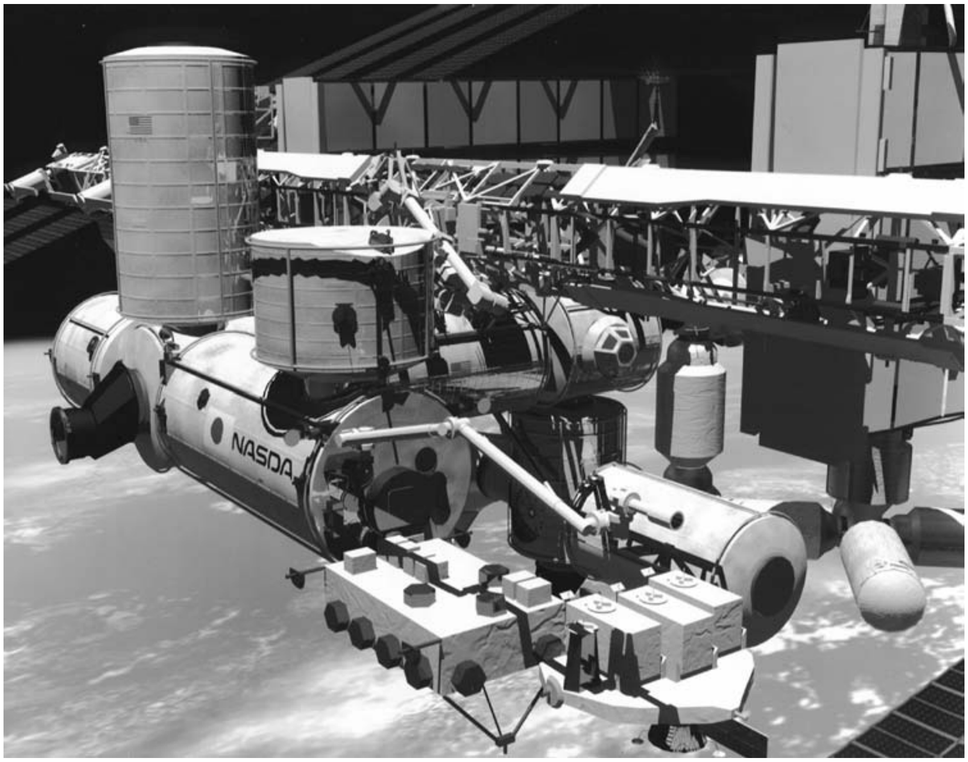


Figure 24. A view of the station showing the Japanese experiment module (JEM), complete with external payloads and remote manipulator arm. Note the windowed cupola module at center, which will be supplied by Europe.
[*Source: Courtesy of NASA.*]

the earth and stars and monitoring of work on and around the station.

There are many technical challenges inherent in the design and construction of such a large structure in space, not least its structural integrity, which despite numerous computer simulations, remained to be proven in flight. Also, because of its size, the probability of impact from space debris will be statistically greater than for previous spacecraft. As a result, the module shells are designed to withstand impacts from objects less than 1 centimeter in diameter without significant damage, while their life support systems should allow at least three minutes for evacuation following a 10 centimeter puncture.

One of the most important questions posed throughout the station's design and development concerned what it would be used for. The answers usually included microgravity research into growing more perfect semiconductor crystals than available on earth; various aspects of biomedical research that could lead to new drugs and procedures for use on earth; and space science applications such as astronomy.

Over and above such scientific endeavors, the ISS is likely also to be used as an in-orbit model shop for future spacecraft technologies; for example, in the testing of large deployable structures such as solar arrays and communications antennas, or for space-qualifying new materials, thermal control hardware and ion propulsion systems, to name but a few. Interestingly, the ISS has already proved itself as a destination for space tourists, the first of which, Dennis Tito, paid a reported US$20 million for a week on the station in 2001. Moreover, the potential for the ISS to act as a departure point for further exploration of the Moon and Mars should not be ignored.

In addition to the engineering challenge of establishing a small community in low-earth orbit, the ISS program has proved to be a political challenge. Agreement on financing and respective responsibilities among international partners has proved difficult and has resulted in modifications and delays. For example, the inclusion of Russia relatively late in the program—following an end to Cold War hostilities—led to a major redesign to incorporate a significant amount of Russian hard-

ware. Although this led to a station in orbit sooner than the U.S. could have produced alone, it did so at the expense of the Russian station, Mir, which was deorbited in March 2001 because Russia could not afford to operate both. Despite the problems, the ISS seems likely to be a model for future international space programs, such as a manned Mars mission, simply because they are unlikely to prove affordable for a single nation.

See also **Space Shuttle; Space Stations, Mir; Space Stations, Skylab**

MARK WILLIAMSON

Further Reading

Covault, C. International Space Station ready for flight. *Aviation Week and Space Technology*, December 8, 1997.

Messerschmid, E. and Bertrand, R. *Space Stations: Systems and Utilization*. Springer Verlag, Berlin, 1999.

Haskell, G. and Rycroft, M., Ed. *International Space Station: The New Space Marketplace*. Kluwer, Dordrecht, 2000.

Smith, M.S. International Space Station cooperation: a view from the US Congress. *Russ. Space Bull.*, 5, 3, 29–30, 1998.

Williamson, M. New star in orbit. *IEE Rev.*, September, 201-205, 1999.

Space Stations, Mir

On June 1, 1970, the USSR launched the Soyuz-9 spacecraft as part of the initial preparation for a reorientation of the Soviet space program in the direction of establishing a permanent human presence in orbit. The Soyuz-9, manned by Andrian Nikolayev and Vitali Sevastyanov, orbited the earth for 18 days. The purpose of the Soyuz flight was to break the record of days spent in orbit, previously set when the crew of the U.S.'s Gemini-7, Frank Borman and James Lovell orbited the earth for 14 days in 1965. While the USSR, like the U.S., had focused its efforts on lunar landings, it now sought to set up an earth-orbiting space station that would serve as a continuing research center and space science laboratory. It was hoped that the research results would aid the Russian economy as well.

Salyut-1, the first of seven satellite stations, was launched on April 19, 1971. It relied on four solar panels for charging its batteries. Salyut-7 was used from 1982 to 1986, and also employed solar panels. It was the "second generation" of the past Salyut models. In 1975, the U.S. and the USSR established the joint Apollo–Soyuz Project. This relationship established through the Apollo–Soyuz Project was the international precursor to the Mir Space Station. The Mir Space Station was designed as a third generation of the Salyut series, utilizing most of the basic original constructs. The first component of the station complex, the Mir Core, was launched on February 20, 1986. The station complex was built in orbit over a period of ten years, with new modules added slowly over time. The Mir Core contained the living quarters and control center. This first module originally had two solar panels, seven computers running off the Strela system, and six ports for the docking of up to two Soyuz crafts. This meant that the station could house up to six people for short periods of time.

The Mir Core had a mass of 20,100 kilograms, was 13.13 meters long, and had a diameter of 4.15 meters. Its three solar panels supplied up to 11 kilowatts of power. Because of the specifications of the Proton rocket used to launch it, the Mir Core was similar in structure to the earlier Salyut craft, and a Proton SL-13 launch vehicle sent the Mir Core into orbit. It contained sleeping compartments, a toilet with a door (the Salyuts had no door for privacy on the toilet), washing station, table, refrigerator, stove, treadmill, and stationary bicycle. Cosmonauts on the space station had a fairly regimented working day. Everyone was required to bicycle 10 kilometers per day and "walk" (they were harnessed to the treadmill to add friction and stability) 5 kilometers per day to prevent muscle and general health deterioration.

In 1987, the Kvant-1 was launched and added to the space station as the second module. The Kvant-1 was 8 meters long, with a pressurized volume of 40 cubic meters. It contained an astrophysics laboratory, but lacked its own solar panels. In December of 1989, the Kvant-2 module was added. The Kvant-2 had a pressurized volume of 60 cubic meters and two solar panels of 50 square meters total area, supplying 7 kilowatts of power. This module, which had its own propulsion system, did not have a singular purpose like the Kvant-1's astrophysics laboratory. The Kvant-2 added six gyrodynes, two tanks for the Rodnick system, two oxygen generators, a shower, a toilet, and myriad other scientific equipment. Kvant-3, or Kristall, provided even more research capabilities when the module was added in June of 1990. Kristall had no gyrodynes, but it did contain two Rodnick tanks, six Ni–Cd batteries, and newly designed solar panels. Kristall had a pressurized volume of 60 cubic meters and its two retractable solar panels, with a combined area of 72 square meters, provided 9 kilowatts of power. In 1995, the Spektr module was added to expand the complex, although this had originally been scheduled for

1991. It had four solar panels and provided more living quarters, mainly used by American cosmonauts. The Spektr had a pressurized volume of 62 cubic meters and 126 square meters of solar panels, which could generate 16 kilowatts of power. In 1996, the final two modules were added, the Priroda module and the Docking module, providing the station complex with remote sensing instruments and a port for a shuttle, respectively. The Priroda module had a pressurized volume of 66 cubic meters. Its earth remote-sensing instruments examined both environmental and ecological changes, such as the spread of industrial pollutants, the height of waves, temperature changes in the ocean, the structure of clouds, and even plankton concentrations.

The last crew left the Mir complex in June of 2000. Up until that point the station had been almost constantly occupied. In total, 43 cosmonauts lived in Mir and 59 others have visited the station for less than one month. On March 23, 2001, the Mir Space Station was brought back to earth and crashed into the South Pacific Ocean. Mir proved man's ability to live in space for extended periods of time, and provided extensive research in numerous fields of study. The Mir provided information on subjects ranging from new galaxies, patterns of ocean and wind direction, and the influence of gravity on biological processes.

See also **Satellites, Environmental Sensing; Space Launch Vehicles; Space Stations, International Space Station; Space Stations, Skylab**

TIMOTHY S. BROWN

Further Reading

Bonnet, R.M. and Manno, V. *International Cooperation in Space*. Harvard University Press, Cambridge, MA, 1994.

Harland, D.M. *The Mir Space Station: A Precursor to Space Colonization*. Wiley, Chichester, 1997.

Oberg, J. *Red Star in Orbit*. New York, Random House, 1981.

Space Stations, Skylab

Skylab was a U.S. space station designed for earth observation, solar astronomy, and research into how humans would work in, and react to, the space environment. It was launched into low-earth orbit—to an altitude of about 440 kilometers—on 14 May 1973 and visited by three groups of three-man crews for periods of 28, 59, and 84 days, until it was vacated in February 1974. Despite later attempts to raise its altitude, Skylab re-entered the atmosphere on 11 July 1979, having completed 34,981 orbits of the earth.

The Skylab program was, in part, an attempt by the National Aeronautics and Space Administration (NASA) to salvage some of the technological investment made in the Apollo lunar program, which was curtailed when government funding was withdrawn. The station was launched by one of the remaining Saturn-V moon rockets and its crews were launched by the smaller Saturn-1B. The station's main section, the orbital workshop (OWS), was made from a converted Saturn-V third stage, effectively replacing propellant and guidance systems with crew accommodation and life-support systems. From an engineering point of view, this produced a welcome but unusual situation in that most of Skylab's structural design had not only been completed, but also flight-tested, before the station left the drawing board.

Attached to the OWS was an airlock module, which allowed spacewalks (to change film in the external cameras and make external repairs) to be conducted; a multiple docking adapter, at which two Apollo spacecraft could be docked; and the Apollo telescope mount (ATM), designed particularly for solar observations (see Figure 25). With a total mass of some 90,000 kilograms and a total habitable volume of about 360 cubic meters, Skylab was by far the largest spacecraft launched at that time; the OWS alone provided a working volume 15 meters long and 6.6 meters in diameter. The contemporary Soviet Salyut stations were about one-third of the size and mass.

The OWS was split by a gridded floor into two sections—a large laboratory area and a smaller wardroom containing sleeping, cooking, eating, and hygiene facilities. Being weightless, the astronauts could attach their specially adapted shoes to either side of the grid to provide stability for certain tasks or simply to avoid floating away. Running along the axis of the OWS was a removable "fireman's pole" which was used for moving between one section and another. Some care was taken in the interior design of the station, particularly the sleeping and eating areas; there were plenty of storage lockers and hardware was color-coded. One of the innovations of Skylab was an enclosed shower, which is something even the current International Space Station lacks.

Skylab was built with two large, deployable solar panels and four smaller ones, having a combined area of 840 square meters and capable of producing about 12 kilowatts of power. Unfortunately, launch vibrations caused the dis-

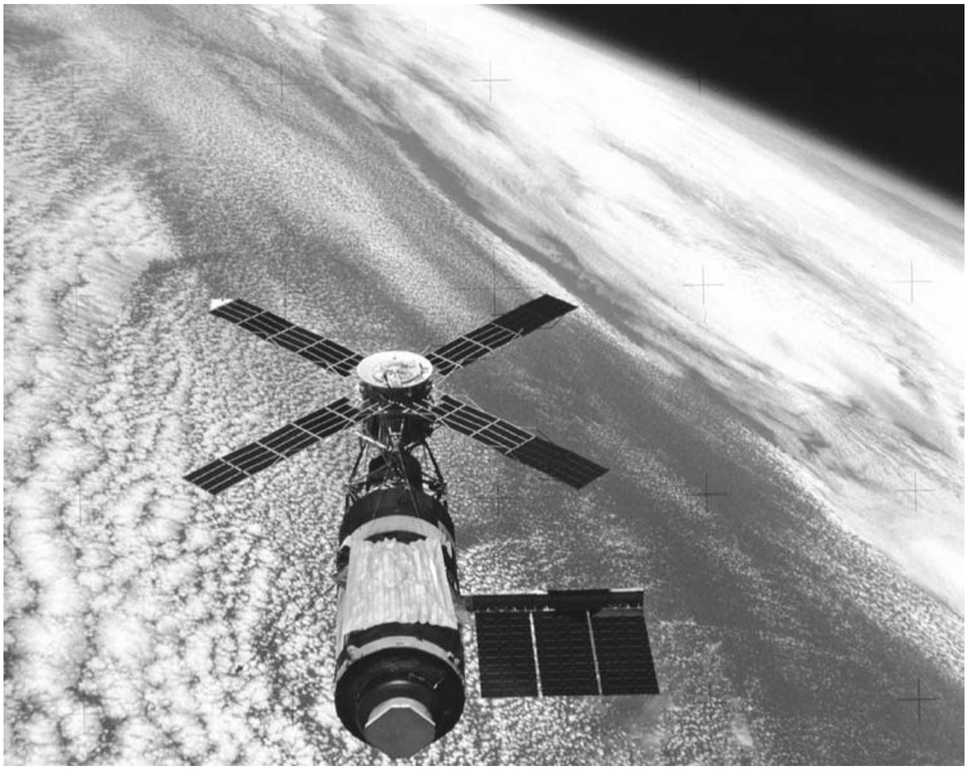


Figure 25. The Skylab space station in orbit, showing a single solar array deployed from the side of the Orbital Workshop (OWS) and four smaller arrays deployed from the Apollo telescope mount (ATM). Note also the concertina-shaped sunshade over the main cylinder, which was erected by spacewalking astronauts to replace thermal insulation torn off during launch.
[*Source: Courtesy of NASA.*]

location of a combined meteoroid–thermal shield on the OWS which tore away one of the large solar arrays and restricted the deployment of the other. Telemetry received from the station shortly after launch indicated that no power was being generated from the arrays and that the internal temperature of the station was increasing rapidly. The first Skylab crew was, however, able to restore the station to an operational condition by conducting a series of spacewalks, deploying a temporary sunshade to replace the missing thermal insulation and releasing the remaining solar panel.

The remainder of the 28-day mission, which was twice the duration of any previous U.S. spaceflight, went according to plan. Despite the time spent on repairs, 46 of 55 planned experiments were completed and the crew suffered no significant ill effects.

Together, the three Skylab crews conducted 270 experiments in the fields of life sciences, astrophysics, solar physics, earth observation, and materials processing, as well as demonstrations of engineering and space technology. The solar observations in particular were highly significant at the time. In total, the program added an impressive 513 man-days to the American manned spaceflight record, a figure not exceeded by the Space Shuttle until its fifth year of operation.

The significance of the Skylab program was that only 12 years after Yuri Gagarin became the first man in space, nine astronauts had proved that people were not only capable of living in space for more than a few days, but could also apply themselves to life-threatening situations and make an otherwise uninhabitable station habitable. Although it proved to be another quarter of a century before the U.S.-led International Space Station would be equally habitable, the experience provided by Skylab was invaluable. Indeed, it was not surpassed until Russia began the in-orbit construction of its Mir Space Station in 1986.

Apart from its scientific and technical success, the program provided the first insights into how people might live together in space. It also showed that even highly trained astronauts were not prepared to adopt the same attitudes on long-

duration missions as they had on previous week-long missions, when every moment was precious and had to be filled with observations and experiments. Where previously, astronauts had tended to keep their criticisms to themselves, in support of good PR and continued public funding, the third Skylab crew opposed aspects of the flight plan and openly criticized the station's amenities. For example, Gerald Carr complained that the soap was like "dog shampoo;" Bill Pogue likened use of the towels, which were made of a fire-resistant synthetic material, to "drying off with padded steel wool;" and Edward Gibson who, along with the others, suffered cold symptoms and nosebleeds (due to high blood pressure in the upper extremities) disliked the station's paper tissues. In addition, they all agreed that the toilet, as characterized by Pogue, had been designed by someone who had never used one. The Skylab program had unintentionally become a psychological study; it had also shown that the human aspects of space station design and operation are as important as the scientific and technical aspects.

See also **Space Exploration, Moon, Manned; Space Stations, International Space Station; Space Stations, Mir**

MARK WILLIAMSON

Further Reading

Arnold, H.J.P., Ed. *Man in Space: An Illustrated History of Spaceflight*. Smithmark, New York, 1993.
Cooper, Jr. H.S.F. *A House in Space*. Angus & Robertson, London, 1977.
Freeman, M. *Challenges of Human Space Exploration*. Springer-Praxis, Berlin, 2000.
Holder, W.G and Siuru, W.D. *Skylab: Pioneer Space Station*. Rand McNally, Chicago, 1974.
Messerschmid, E. and Bertrand, R. *Space Stations: Systems and Utilization*. Springer Verlag, Berlin, 1999.

Useful Websites

Benson, C. and Compton, W. *Living and Working in Space: A History of Skylab*: http://history.nasa.gov/SP-4208/contents.htm

Spark Transmitters, *see* **Radio Transmitters, Early**

Spectroscopy and Spectrochemistry, Visible and Ultraviolet

Spectrum analysis was launched in 1859–1860 by the physicist Gustav Robert Kirchhoff and the chemist Robert Wilhelm Bunsen. They demonstrated experimentally that each chemical element has its own characteristic set of spectrum lines, which it emits or absorbs when heated to the state of a radiating gas by a Bunsen-burner flame, an electric arc or a spark. Within the context of nineteenth century science, the various patterns in these spectra could only be described, catalogued and mapped—and quite extensively so. An understanding of these patterns of series, bands and the splitting of these lines into components by an electric or magnetic field (Stark and Zeeman effect) had to wait until the twentieth century. Niels Bohr's atomic model of 1913 interpreted spectrum lines as the result of electron jumps between stable orbits around the nucleus. Quantum mechanics abandoned orbits but kept the notion of energy levels henceforth to be calculated on the basis of Schrödinger's equation, leading to an even better agreement between observations and its theoretical predictions concerning the transition terms (selection rules) and energies (line frequencies).

The dramatic success of spectrum analysis after 1860 caused this qualitative analytic technique to make quick inroads into the chemist's or pharmacist's laboratory, the astronomer's observatory, the physician's hospital, and even the judge's court. The most notable use of the new technique involved identifying the presence of the various elements in a given sample. It led to the surprising finding that the metal lithium, for instance, hitherto considered quite rare, was among the more ubiquitous chemical elements. News that it was possible to decipher the chemical composition of the sun and ultimately the stars spread fast: Who could fail to be impressed by the fact that only microscopic amounts of sodium (3×10^{-9} g)—a mere teaspoon of salt in a full swimming pool—were needed for detection of its characteristic yellow D lines in a Bunsen flame? Using a simple pocket spectroscope, a steel caster could now easily identify the exact instant of decarbonization of molten steel, the moment just before it loses its fluidity.

Quantitative emission spectroscopy had a much more difficult start, however. Around 1906, Comte Arnaud de Gramont started to record methodically the detectability of characteristic lines that remain visible as long as the slightest trace of a substance is present. These *raies ultimes*—ultimate or residuary lines—were the most reliable indicators of the respective chemical elements in a sample. For 83 different elements including (besides most metals) a great many rare earths and nonconducting elements, de Gramont listed the low number of 307 ultimate and penultimate lines. As a regular analyst for four French steel mills, de Gramont

helped improve their production significantly because he could readily report to them the presence and approximate concentrations of aluminum, boron, cobalt, chromium, copper, manganese, molybdenum, nickel, silicon, titanium, vanadium, and tungsten in their steel. During World War I, de Gramont and a few assistants used the method in a broad array of military applications. Among them were quick and efficient examinations of the structural frames or valves of zeppelins, shrapnel from long-range guns, or ignitors of aircraft.

The spectroscopic laboratory of the National Bureau of Standards in Washington was among the very first to adopt de Gramont's method of "practical spectrographic analysis." After the war, William F. Meggers applied it to the chemical analysis and quality control of noble metals, such as gold and platinum, for the U.S. Mint in San Francisco. Promising analytical and metallurgical applications were explored by the American Brass Company, in Waterbury, Connecticut, as well as by a few other U.S. industrial laboratories. Despite the obvious importance of de Gramont's work for the French war machine, strangely enough, Germany did not implement anything even remotely similar. Carl Friedrich (called Fritz) Löwe was one of the earliest active promoters of its industrial applications among chemists and physicians. Löwe's frankness about the missed opportunities during the recent—lost—World War was effective in arousing renewed interest in the method. His touting of instruments by the Zeiss Company in Jena reached German-speaking audiences. Frank Twyman fulfilled a similar promotional function in the Anglo-Saxon world for his company, Adam Hilger Ltd. in London, and the Stockholm professor of experimental biology Henrik Gunnar Lundegårdh pointed out many applications in mineralogy, biochemistry, plant physiology, and agricultural chemistry. By 1930, a typical spectrochemical procedure took no longer than 20 minutes (including development of the photographic plate). A decade later, the industrial pressure for ever-higher production rates had "super-speed analysts"—as Meggers called them—reduce this time to a minute or two. The laboratories of Ford Motor Company, for instance, carried out large numbers of analyses at high speed: samples were sent by pneumatic tube from the foundry to the spectrographic laboratory and just a few minutes after receipt of the sample, the results were available back on the factory floor. This method left the sample virtually unscathed and allowed close examination of local differences of parts of the surface or various layers of it. By contrast, wet chemical analysis inevitably yielded average results because the sample had to be analyzed in solution. Other applications of spectrochemical analysis after 1930 include:

- Absorption spectrophotometry of organic solutions for identification of hormones, vitamins, and other complicated substances.
- Testing for silver or boron content in the mining industry.
- Routine quality control in the metallurgical and chemical industries, including monitoring of isolation or separation processes.
- Soil analysis for agriculture and plant physiology.
- Applications in the food packing industry (e.g., checking the dissolving rate of inner coating of a tin can by measuring two or three parts of aluminum or lead per ten million, or testing chocolate or chewing gum wrappers, and whiskey distilling vats).
- Forensic analyses or autopsies for detection of trace amounts of toxins (e.g., thallium from rat poison, which is ascertainable in hair samples).
- Analysis of fusible alloys of tin for safety valves or fire sprinklers (to trace impurities such as lead and zinc, which may raise the melting point by undesirable amounts if present in proportions of as little as one part in ten thousand).
- Archaeometric comparisons of the precise composition of metals and alloys from various locations (sometimes enabling archeologists to infer where a certain piece had been manufactured, or conclusions about the geographic and temporal spread of certain technologies or skills).
- Plentiful applications in mineralogical analysis (which, as we have seen in the case of de Gramont, had initiated some of the earliest efforts in quantitative spectrochemical analysis).

The plethora of possibilities turned spectrochemistry into a vibrant and popular field. The industrial world embraced it in the following decades, setting up thousands of spectrochemical laboratories. The boom in this field of research can be gauged somewhat by publication statistics in spectrochemistry: 1467 books and papers, and half a dozen treatises were indexed in the first part of Meggers' and Scribner's bibliographic survey, covering 1920–1939. A total of 1044 contributions were made in the short period of

World War II, 1940–1945, another 1264 in the next five postwar years, and 1866 in the period 1951–1955. A true explosion in the literature followed, with an exponential growth in many scientific fields leading to an estimated total of 10,000 spectrochemical publications by 1963.

Visual resources like atlases were an integrated part of the effective marketing strategy of the major spectrograph manufacturers: Zeiss, Hilger, or Fuess. The most ambitious inventorization effort was the famous Massachusetts Institute of Technology (MIT) table of 100,000 wavelengths. It was compiled with specially developed spectrophotometers capable of automatically measuring, computing, and recording the wavelengths of spectrum lines, thus speeding up these operations some 200-fold. As one of the major teaching and research centers for spectroscopy, the MIT started to host annual summer conferences on spectroscopy in 1933. An initial attendance of 69 persons in the first year increased to 233 in 1938, 250 in 1939, and 302 in 1942. The series was interrupted for the remaining war years but resumed thereafter. The rapidly expanding market for spectrographs and spectrometers led to the initiation of specialized events such as the National Instrument Conference and Exhibit. An overlapping interest in spectrochemical instrumentation and techniques motivated the Society for Analytical Chemistry of Pittsburgh (SACP, founded in 1943) and the Spectroscopy Society of Pittsburgh (SSP, founded in 1946) to combine their annual meetings in 1949. The joint meetings of these hitherto moderately sized societies, held every March since 1950 under the acronym Pittcon (Pittsburgh Conference and Exposition on Analytical Chemistry and Applied Spectroscopy), transformed spectrochemistry to the point that the convention eventually outgrew this steel-producing city and its organizers were forced to find other locations. Whereas the first Pittsburgh Conference offered 56 presentations and 14 exhibits by commercial instrument makers, the 1990 conference (held at the Jacob Javits Center in New York) coordinated more than 1200 talks and 25 symposia, over 3000 instrument exhibits by over 800 commercial instrument makers, and 12,500 hotel bookings.

Both the high demand for spectrochemical techniques during World War II and the ubiquitous pressure for ever-faster results led inevitably to increased substitution of quasi-instantaneous photoelectric detection in photographic recording. This elimination of photographic development and densitometry in favor of photomultipliers and electronic automation was pushed particularly in the U.S. in companies like Dow Chemical Company in Midland, Michigan, Perkin Elmer in Boston, Baird Associates (BA) in Cambridge, Massachusetts, Applied Research Laboratories (ARL) in Glendale, California, and National Technical Laboratories, renamed Beckman Instruments in 1950, whose direct-reading spectrometers flooded the international market in the 1950s. Advertisers claimed these ''analysis automats'' made ''all routine spectrochemical analyses with dispatch and precision,'' and in the 1950s and 1960s they eventually did. With these improvements also came a rapid expansion of potential applications, especially in infrared spectroscopy.

The near-ultraviolet (UV) had already been explored photographically by Eleuthère Mascart and Alfred Cornu in the nineteenth century. However, glass optics absorb radiation past 3440 Å (1 angstrom = 10^{-10} meters), which could be circumvented by using quartz or Iceland spar prisms; ozone absorbs wavelengths past 2900 Å; and the gelatin emulsions of photographic plates those past 1850 Å. Therefore further progress had to await the development of high-vacuum spectrographs and gelatin-free emulsions. The latter two fields were pioneered by Victor Schumann in Leipzig, who reached wavelengths down to approximately 1000 Å, and Theodore Lyman at Harvard University who discovered the ultraviolet series of hydrogen in 1914. After World War II, grating spectrographs were mounted on rockets and propelled out of the terrestrial atmosphere to record high-resolution solar UV-spectra.

See also **Iron and Steel Manufacture; Materials and Industrial Processes; Spectroscopy, Raman; Spectroscopy, X-Ray Fluorescence**

KLAUS HENTSCHEL

Further Reading

Brand, J.D. *Lines of Light. The Sources of Dispersive Spectroscopy, 1800–1930*. Gordon & Breach, Amsterdam, 1995. Mostly covers the development of theoretical interpretations including quantum mechanics.

Hentschel, K. *Mapping the Spectrum. Techniques of Visual Representation in Research and Teaching*. Oxford University Press, 2002. Mostly covers experiments and applications and further references to primary literature given therein.

Spectroscopy, Infrared

The investigation of invisible optical radiation gained increasing attention during the second half of the nineteenth century, notably with the identi-

fication by William Abney of infrared spectra characteristic of different chemical compounds in 1882. Infrared spectroscopy nevertheless grew to become a popular analytical technique only during the mid-twentieth century.

Infrared spectroscopy maps the absorption, emission, or reflection of radiation as a function of wavelength. The infrared spectrum of a chemical absorbing compound constitutes a "fingerprint" which identifies functional groups in its molecules, and can reveal a wide range of physical and chemical properties of matter. Despite such potential utility, its study at the turn of the twentieth century appeared highly unpromising: development was constrained by the very weak signal produced by infrared radiation. Infrared energy passing through a spectrometer causes temperature changes that are typically less than a few thousandths of a degree. The corresponding electrical signal using available detectors was swamped by other influences.

State-of-the-art detection during the 1920s combined an infrared detector with a mirror galvanometer. However, this direct-current signal varied not only with the weak infrared contribution, but with the temperature of the surroundings. To obviate the need for constant zero-adjustment of the galvanometer, in 1929 A. Pfund devised a "resonance radiometer" which detected "chopped" (mechanically interrupted) radiation using a mechanically tuned mirror galvanometer. This modulation technique was a key feature of subsequent instruments. Electronic components developed originally for radio communication began to be applied to instrumentation amplifiers by the late 1920s, and Pfund's method was extended by a capacitively coupled amplifier. Its successors relied solely on electronic components to yield a high-gain tuned amplifier, notably the stable alternating current (AC) amplifier for thermopiles (essentially a compact combination of blackened thermocouples in series) developed by L. C. Roess. As Pfund and Firestone had done, Roess modulated the light beam by interrupting it periodically with a chopper disk, and tuned the amplifier to the modulation frequency to reinforce the very weak infrared signal preferentially. Combined with an automatic pen recorder, the Roess amplifier was quickly adopted by spectroscopic researchers for use with the new thermistor bolometers (blackened semiconductor devices having temperature-dependent resistance and wide spectral response) invented in 1946.

Like spectrometers for the visible portion of the spectrum, the first infrared spectrometers employed either prisms or diffraction gratings as the dispersive elements. Such prisms, fabricated from alkali salts transparent to infrared radiation, were combined with a thermopile detector to extend observations from the visible portion of the spectrum (about 0.7 micrometers) to 40 micrometers. Through the 1930s and early 1940s, designs proliferated as spectrometers were custom-built by laboratory-based researchers. During the World War II, infrared spectroscopy proved valuable for analyzing and quantifying chemical constituents and for monitoring the production of essential materials such as petroleum products and synthetic rubber.

Automatic spectrophotometers for the visible spectrum had been commercialized in the mid-1930s, and infrared versions began to appear increasingly after the war. With the refinement of servo-mechanisms during this period, more spectrometers began to incorporate automatic recording. Increased automation considerably eased the operator's burden of setting amplifier offsets and gains, adjusting the width of slits, and altering the rate of scanning the wavelength to compensate for the dramatic variations in energy across the spectrum.

By the early 1950s there were over a dozen manufacturers of infrared spectrometers, ranging from desk-size research-grade instruments to simpler bench-top spectrometers. Prominent manufacturers during this period included Perkin-Elmer, Grubb-Parsons, Beckman, and Leitz. The market for such instruments centered on chemical research laboratories. The spectrometers were relatively costly and demanded a stable room temperature and knowledgeable operators.

Using such instruments, infrared spectroscopy became increasingly routine at ever-longer wavelengths. Given the low sensitivity of available detectors and the relatively weak infrared sources of radiation, however, there was a limit to such extension. By the late 1960s the best spectrometers, using highly automated control mechanisms and multiple diffraction gratings for different spectral regions, could extend measurement to a wavelength of about 300 micrometers.

To investigate such energy-starved regions of the spectrum and naturally weak emitters such as astronomical sources, a new instrument principle was developed. So-called "Fourier spectroscopy" proved considerably more efficient than conventional methods. Fourier spectroscopy does not disperse the radiation from the light source into separate spectral components, which are then individually measured; instead, the radiation passes through an interferometer, which modulates the radiation and passes it to the infrared detector.

An interferometer consists, in its simplest form, merely of a beamsplitter and two mirrors (see Figure 26). The beamsplitter divides the incoming radiation into two parts. One part passes directly to a fixed mirror, which reflects it back to the beamsplitter. The second part of the radiation passes to a similar mirror, which can be moved backwards to lengthen the optical path. It, too, reflects its beam of light back to the beamsplitter, which recombines the two components to yield an output beam that passes to the optical detector. By moving the adjustable mirror, the light traveling along that arm of the interferometer is successively delayed with respect to that of the other arm. Different wavelengths comprising the radiation are brought into and out of step, interfering and thus changing the intensity of the combined beam. The intensity of the combined beam, modulated as a function of mirror position, is the Fourier transformation of the optical spectrum.

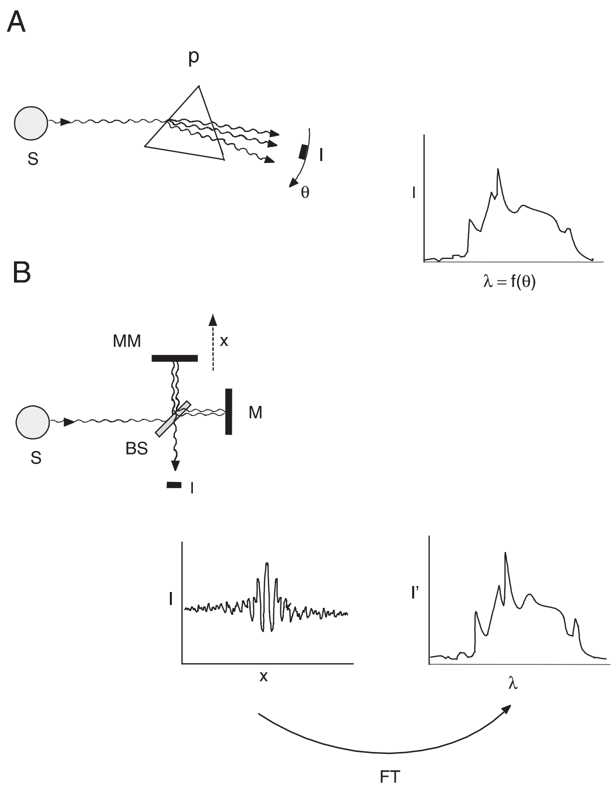


Figure 26. Distinguishing features of infrared spectroscopies. (A): Dispersive infrared spectroscopy: *S*, infrared source; *P*, prism or diffraction grating; *I*, intensity measured by detector swept through angle $\tilde{0}$. In the resulting spectrum I(λ), the wavelength λ is a function of θ. (B): Fourier spectroscopy: *S*, infrared source; *BS*, beam-splitting mirror; *M*, fixed mirror; *MM*, moving mirror, translated through distance *x*; *I*, intensity measured by detector. The resulting record of *I* versus *x*, or "interferogram," is related to the spectrum $I'(\lambda)$ (or, more accurately, $I'(1/\lambda)$) by the Fourier transformation *FT*. Both techniques involve specific complications to calibrate the wavelength scale and instrumental response.

Physicists found three distinct advantages of Fourier spectrometers compared to their dispersive predecessors: (1) the replacement of narrow slits with a full optical aperture provides higher optical throughput, or *etendué* (Jacquinot, 1954); (2) the measurement of all wavelengths simultaneously rather than sequentially improves signal quality, or alternatively can reduce measurement time (Fellgett, 1958); and (3) the precise determination of mirror position via a laser reference wavelength provides spectra with much better wavelength precision (Connes, 1966). In combination, these advantages make the technique considerably more sensitive optically and amenable to more extensive data analysis than dispersive techniques.

However, the instrumentation was attractive only to narrow audiences during its first two decades. The operating principles of Fourier spectrometers proved considerably less intuitive than their dispersive counterparts, especially for chemists. Moreover, the Fourier transformation, which converted the modulated signal into an infrared spectrum, demanded expensive and all-too-slow computers. Despite the dramatic speed increase provided by the fast Fourier transform (FFT) from 1966, Fourier spectroscopy made inroads only with isolated groups of physicists. By the mid-1970s, computer power and cost had improved sufficiently to allow commercially packaged and highly automated Fourier spectrometers, rechristened "FTIR" (Fourier transform infrared) instruments, to overtake dispersive instruments in the chemistry market. This process was essentially complete by the late 1980s.

Modern infrared spectroscopy is based principally on Fourier spectrometers coupled to increasingly powerful computers for identification and quantification. Infrared analysis is one of the most commonly used techniques in analytical chemistry (e.g., environmental monitoring), chemical process control, and remote sensing.

See also **Infrared Detectors; Spectroscopy and Spectrochemistry, Visible and Ultraviolet; Spectroscopy, Raman; Spectroscopy, X-Ray Fluorescence**

SEAN F. JOHNSTON

Further Reading

Abney, W.W. and Festing, E.R. On the influence of the molecular grouping in organic bodies on their absorp-

tion in the infra-red region of the spectrum. *Proc. R. Soc. London*, 31, 416, 1882.
Chantry, G.W. *Submillimetre Spectroscopy*. Academic Press, London, 1971.
Connes, J. and Pierre. Near-infrared planetary spectra by Fourier spectroscopy. *J Optic. Soc. Am.*, 56, 896, 1966.
Fellgett, P. A contribution to the theory of the multiplex spectrometer. *J Phys. Radium*, 19, 187, 1958.
Firestone, F.A. A periodic radiometer for eliminating drifts. *Rev. Sci. Instrum*, 3, 163, 1932.
Griffiths, P.R. and de Haseth, J.A. *Fourier Transform Infrared Spectrometry*. Wiley, New York, 1986.
Hardy, A.C. History of the design of the recording spectrophotometer. *J. Optic. Soc. Am.*, 28, 360, 1938.
Jacquinot, P. The luminosity of spectrometers with prisms, gratings or Fabry-Pérot etalons. *J. Optic. Soc. Am.*, 44, 761, 1954.
Johnston, S.F. *Fourier Transform Infrared: A Constantly Evolving Technology*. Ellis Horwood, Chichester, 1991.
Johnston, S.F. In search of space: Fourier spectroscopy 1950–1970, in *Instrumentation: Between Science, State and Industry*, Joerges, B. and Shinn, T., Eds. Kluwer, Dordrecht, 2000.
Kendall, D.N., Ed. *Applied Infrared Spectroscopy*. Reinhold, New York, 1966.
Michelson, A.A. On the application of interference-methods to spectroscopic measurements—I. *Philosoph. Mag.*, 31, 338, 1891.
Roess, L. C. Vacuum tube amplifier for measuring very small alternating voltages. *Rev. Sci. Instrum.*, 16, 172, 1945.
Strong, J. Apparatus for spectroscopic studies in the intermediate infrared region—20 to 40 μm. *Rev. Sci. Instrum.*, 3, 810, 1932.

Spectroscopy, Raman

When Sir C. V. Raman discovered the spectroscopic technique that bears his name in 1928, it was the result of experiments conducted with inexpensive equipment at the Indian Association for the Cultivation of Science (IACS) in Calcutta, India, far removed from the acclaimed laboratories of the Western world. Ironically, the idea for the experiments came during the return trip from his first visit to London in 1921.

Intrigued with the deep blue color of the Mediterranean Sea, Raman was skeptical of Lord Rayleigh's explanation that the color of the sea was simply the reflection of the color of the sky. Formulating his thoughts while still on board the *SS Narkunda*, Raman dashed off a letter to the editors of *Nature* as soon as he reached Bombay. Within a short time, he was able to prove that the ocean's color was not a mirror of the sky above, but the result of sunlight being scattered by the water molecules.

Analyzing the light scattered by a liquid is difficult, and Raman's early work used visual observations of color changes rather than determining the exact wavelength of the light associated with the color, as shown in Figure 27.

Raman's experiment of discovery is outlined in Figure 28. Using a violet filter to isolate the violet portion of sunlight, Raman looked at the scattered light that emerged at right angles to the original beam after it had passed through a liquid sample. Most of the light emerged unchanged as the original violet color. However, when Raman and K.S. Krishnan used a green filter to intercept the scattered light, they found that the spectrum of the scattered light had a weak green component as well as violet, meaning that the wavelength had been shifted. Raman quickly found some 60 samples that exhibited this effect, which soon became known as the "Raman effect."

In spite of this initial success, Raman's problem was that the effect was very weak, as only one in a million of the scattered photons actually changed wavelength. Although sunlight was plentiful in Calcutta, its intensity was barely adequate. In 1927 the IACS obtained a refracting telescope that Raman used to focus the sunlight into a more powerful source. By 1928, the newly available mercury-arc lamp replaced sunlight, and a quartz spectroscope replaced the visual measurement of color. Raman was now able to make precise measurements of the wavelengths of the scattered light. His refined quantitative results had an immediate impact on the scientific community, and resulted in the award of the Nobel Prize in Physics just two years later, in 1930.

At first, the Raman effect found its niche in physics. Seeking to explain the origin of the Raman effect, physicists were unavoidably drawn to using it

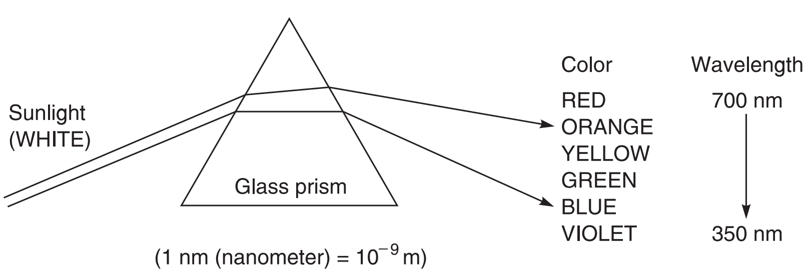


Figure 27. Analyzing light scattered by a liquid using visual observations of color changes.
[*Courtesy of the American Chemical Society*.]

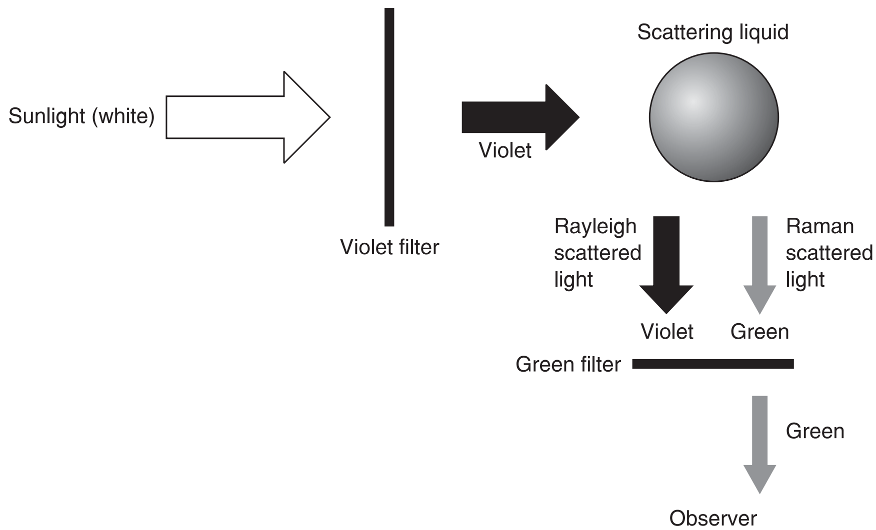


Figure 28. C. V. Raman's experiment of discovery. [*Courtesy of the American Chemical Society*.]

in their study of the vibration and rotation of molecules and ultimately the structure of those molecules. According to R.W. Wood of Johns Hopkins University, Raman's study of light scattering was "one of the most convincing proofs of the quantum theory," a revolutionary new concept that at the time was drastically changing the way that physicists thought about the structure of matter.

After seven years, the intense interest in the Raman effect by physicists waned, but chemists then found that it could be used as an analytical tool because of several characteristics. The Raman effect used by physicists thus became Raman spectroscopy when used by chemists. Since the wavelength of spectral lines of the scattered light emitted by a particular substance was independent of the incident light but characteristic to the sample (related to the internal vibrational motion of the molecule), the lines served as a fingerprint for identification. Moreover, the intensity of those lines was an indication of the amount of the substance in a sample, providing both a qualitative and quantitative determination in a single analysis. Unlike most other analytical techniques then in use, Raman spectroscopy could be applied to solids, liquids or gases, and even aqueous solutions, which had always been difficult to analyze.

Raman spectroscopy fell into disuse after World War II, unable to compete with infrared spectroscopy, a new analytical tool that had been developed during the war. Because infrared spectrometers used sensitive electronic detectors that also came out of wartime research, their operation required little training and became quite routine. By comparison, Raman spectroscopy still needed darkroom facilities and highly trained operators, and attention naturally turned to the newer and easier method.

In the 1960s, however, chemists again became interested in Raman spectroscopy because of a new development that came from physics—the laser. Many years earlier, Raman had constantly sought new and more intense light sources to enhance the scattered light and its subsequent measurement. The laser provided an intensity that far surpassed any previous light source; it also produced quasi-monochromatic light (narrow band of wavelengths), enabling the shifted wavelengths to be distinguished from the source wavelength. When the laser was coupled with Fourier transform techniques and new computers that could quickly process data, a new period of Raman spectroscopy opened in the 1980s when chemists began using the new models that were now commercially available. Today, a Raman spectroscopy system typically includes a laser source (filtered for monochromaticity), focused on a sample; the scattered light is usually detected by a multichannel charge-coupled device array (CCD), which can read the whole scattered spectrum (different wavelengths) simultaneously. Prior to this, point detectors were consecutively positioned at each point in the spectrum.

Raman spectroscopy is used in a myriad of applications, and new ones continue to be found. Micro-Raman spectroscopy is used as an analytical tool for the investigation of objects of art, antiquities, archaeological remains, and other valuables, and has been used to study the embalming techniques of ancient Egyptians. Raman spectroscopy has the potential for rapidly determining the presence of pathogenic bacteria, and will be used to analyze the mineral composition of the Martian

surface. Other uses of Raman spectroscopy range from monitoring industrial manufacturing processes in the petroleum and pharmaceutical industries to analyzing illegal drug samples without destroying the evidence seal on the sample bag taken at the crime scene. Nuclear waste materials can be analyzed at a safe distance and paint can be studied while it is drying to improve its adherence properties. Fabric dyeing, foods, polymers, semiconductors, minerals, and medicine are just a few of the other areas currently using Raman spectroscopy. According to Kathy Kincade, Raman spectroscopy even "has the ability to provide specific biochemical information that may foreshadow the onset of cancer and other life-threatening illnesses."

When he described his "new kind of radiation of light emission from atoms and molecules" in 1928, Raman concluded, "We are obviously only on the fringes of a fascinating new region of experimental research which promises to throw new light on diverse problems… It all remains to be worked out." What has been "worked out" since then was recognized by the American Chemical Society in 1998, when it designated the Raman effect as an International Historic Chemical Landmark.

See also **Spectroscopy and Spectrochemistry, Visible and Ultraviolet; Spectroscopy. Infrared; Spectroscopy, X-Ray Fluorescence**

JAMES J. BOHNING

Further Reading

McCreery, R.L. *Raman Spectroscopy for Chemical Analysis*. Wiley, New York, 2000.
Bohning, J.J. *The Raman Effect*. American Chemical Society, Washington D.C., 1998.
Hibben, J.H. *The Raman Effect and Its Chemical Applications*. Reinhold, New York, 1939.
Tu, A.T. *Raman Spectroscopy in Biology*. Wiley, New York, 1982.
Venkataraman, G. *Journey Into Light*. Penguin, New Delhi, 1994.
Kincade, K. Raman spectroscopy enhances *in vivo* diagnosis. *Laser Focus World*, July, 83–91, 1998.
Various. *Nobel Lectures (Physics)*. Elsevier, New York, 1965, 263–277.
Raman, C.V. *The New Physics*. Philosophical Library, New York, 1951.

Useful websites

Aspects of Raman Spectroscopy: http://www.spectroscopynow.com/
Ghent University Raman Group and the Analysis of Artworks: http://allserv.rug.ac.be/~lmoens/Raman/Welcomeraman.html
Raman Research Institute: http://www.rri.res.in/
Nobel Prize in Physics 1930: http://www.nobel.se/physics/laureates/1930/

Spectroscopy, X-Ray Fluorescence

X-ray fluorescence (XRF) spectroscopy is a widely used method for identifying the chemical composition of almost any material, regardless of its quantity or form. The analysis is useful for two primary reasons: it is noninvasive and nondestructive to the sample material; it yields an easily identifiable and reliable atomic composition of materials, excluding only those primarily constituted of elements lighter than aluminum.

The history of XRF spectroscopy begins with Wilhelm K. Roentgen's 1895 discovery of x-rays, which earned him the 1901 Nobel Prize. Less than ten years later, Charles G. Barkla discovered a relationship between the x-rays emitted from a sample material and that sample material's atomic weight. Then in 1913, Henry Gwyn Jeffreys Moseley renumbered the periodic table of elements by measuring the x-ray fluorescence of each element. Prior to his reorganization, the periodic table was arranged by atomic weight rather than atomic number.

Moseley is also credited with inventing the first XRF spectrometer, which used electrons as a rather inefficient energy source. Other scientists continued to build variations of x-ray detectors with little progress in efficiency until 1948 when Herbert Friedman and LaVerne S. Birks built an XRF spectrometer with a highly sensitive x-ray detector that utilized a Geiger Counter. Their innovation increased the number of elements that could be identified. Then in the 1960s and 1970s, computers could rapidly analyze data, allowing for readings of multiple elements at a time and larger quantities of sample materials to be tested. Widespread usage of XRF spectrometers increased as they become more and more efficient and even portable. In fact, the National Aeronautics and Space Administration (NASA) sent XRF spectrometers on both the Apollo-15 and -16 missions, as well as to investigate the asteroid Eros, and to Mars onboard the Pathfinder in 1997 and the Rover in 2004.

XRF spectroscopy is a relatively accessible process for data collection. To conduct research, one needs a sample, an energy source, an x-ray tube, an x-ray detector, and a computer for data analysis. As noted above, one of the advantages of XRF spectroscopy is that nearly any kind of unknown material may be identified, whether it is in solid, gaseous, or liquid form. The choice of

energy source is largely dependent on the form of the sample material and the particular application. Energy sources vary from x-rays to alpha-particles to beams of electrons. Via an x-ray tube, the energy source bombards the sample material. An x-ray detector or x-ray fluorescence spectrometer then measures the fluorescent light emitted from the sample atoms. Computers then are used to compute the identity of the element and its quantity within the sample. One disadvantage to this process is that most energy sources for XRF spectrometers are radioactive and must be replaced on a regular basis due to the normal decay of radioactive materials.

The basic process of XRF spectroscopy begins when, depending on the particular application, the sample is initially bombarded through an x-ray tube with an energy source such as x-rays, alpha-particles, or beams of high-energy electrons. When electrons in the innermost orbital shells of sample atoms absorb the energy, they are ejected from the atom. When this happens, electrons from the higher-energy outer shells of the atom move inward to fill vacancies in the lower-energy orbitals to stabilize its atomic structure. The movement of the electrons from higher-energy orbitals to lower-energy orbitals causes energy to be emitted in the form of fluorescent light. This fluorescent light, or x-ray, is the characteristic signature of that particular element.

The x-rays are reliable identifiers of the element because of their relationship with other properties of atomic physics. The energy emitted by the transitioning electron will be equal to the difference between the binding energies of the two orbitals occupied by that electron. Because the difference of two specific orbital shells of a given element is constant, the energy emitted by a transitioning electron is also constant and is therefore characteristic of that element. Thus, XRF spectroscopy is used to qualitatively identify the elemental content of a sample material.

XRF spectroscopy can be used to quantitatively measure the amount of a given element within a sample material. Scientists can assess the count rate or peak intensity of the wavelength of the fluorescent x-ray emitted by the transitioning electron. Specifically, the count rate refers to the number of emitted fluorescent photons per unit of time. The count rate can then be used to establish the quantity of a particular element within the sample. Analysis of count rates is easily accomplished with the help of computers.

Applications from the evaluative processes of x-ray fluorescence spectroscopy touch the everyday lives of people everywhere. Portable versions of XRF spectrometers are widely used for field applications, whether the site is an ancient city long buried by volcanic activity or the surface of Mars. The primary industrial use of XRF spectrometry is for quality control of material composition, from raw forms to finished products. Examples include NASA's use of XRF spectroscopy to evaluate the geology of Mars during the 2004 Rover expedition, and museums' use of the process to identify pigments in rare paintings for purposes of restoration. The U.S. Food and Drug Administration evaluates the content of vitamins and other drugs with XRF spectroscopy while archeologists use XRF spectrometers to identify and date artifacts by their mineral contents.

In just over one hundred years since the discovery of x-rays, XRF spectroscopy has become one of the most reliable and widely used methods for chemical composition analysis. Biologists, chemists, museum curators, health inspectors, forensic scientists, medical doctors, ecologists, mineralogists, archeologists, and many university students all use XRF spectroscopy to identify elemental constituents of a vast range of materials. The process is vital for quality control of raw materials and finished products within numerous industrial settings. By traveling into space to classify yet-to-be explored worlds as well as back in time to identify ancient artifacts, scientists use XRF spectrometers to expand the domain of knowledge.

See also **Spectroscopy and Spectrochemistry, Visible and Ultraviolet**

MARY C. INGRAM

Further Reading

Jenkins, R., Ed. *X-ray Fluorescence Spectrometry*, 2nd edn. Wiley, New York, 1999.

Rath, B.B. and DeYoung, D.J. The Naval Research Laboratory: 75 Years of Materials Innovations. *JOM: The Member Journal of The Minerals, Metals, and Materials Society*, 50, 7, 14–19, 1998.

Scott, D.A. The application of scanning x-ray fluorescence microanalysis in the examination of cultural materials. *Archaeometry*, 44, 475–482, 2002.

Van Grieken, R.E. and Markowicz, A.A., Eds. *Handbook of X-ray Spectrometry*. Marcel Dekker, New York, 1992.

Useful Websites

Mars Pathfinder Instrument Descriptions: http://mpfwww.jpl.nasa.gov/mpf/sci_desc.html

NEAR: Near Earth Asteroid Rendezvous Mission: http://near.jhuapl.edu

Sports Science and Technology

The discipline of sports science in medicine and technology is a combination of biomechanics, kinesiology, and anatomy. Within these parts, sports science dates back to Aristotle and Archimedes, when sport was an integral part of Roman and Greek life and inspired a fascination with the mechanics of the human body. By the twentieth century, sports science also began to include varying disciplines such as engineering, polymer science, psychology, and psychiatry, and performance-enhancing pharmaceuticals inspired by high-stakes competition. Today, the business of sports equipment—from running shoes to stair machines—is a multibillion dollar industry based on the continuing evolution of sports science.

As a regular to Gladiator competitions and a physician to the athletes, the second century AD Greek Galen laid out the basic motor functions of the human body. He identified the processes of muscle contraction and the influence of the mind—or "animal spirits" as he called it—on the performance of the body. Isaac Newton's basic theories of scientific reasoning, specifically the study of modern dynamics, laid the groundwork for understanding the relationship between forces and their effects, based on his laws of rest and movement published in *Principia Mathematica Philosophiae Naturalis*. Not until Leonardo da Vinci studied the structure of the full human body were the specific mechanics necessary for simple movements grasped—from walking and sitting to jumping and sprinting. From that time, interested physicians, engineers, and scientists have pursued elements of biomechanics and kinesiology. By 1865, Gullaume Benjamin Amand Duchenne published *Physiologie des Mouvements*, which identified each individual muscle in relation to its range of movement, making it the basis of most sports science work today. Twenty years later, Etienne-Jules Marey studied the motion of humans and animals through photography, publishing his groundbreaking work, *Le Mouvement*, which examined, frame by photographic frame, the intricate interaction of nerves, muscle, and bone.

With the onset of World War I, an interest in biomechanics swelled, though under unfortunate and sometimes inhumane circumstances: Analysis of soldiers on the march and of prosthesis for amputees and the exhaustive study of prisoners of war and concentration camp prisoners all furthered the understanding of biomechanics and kinesiology. In particular, Frenchman Jules Amar, evaluated the human gait and the task performances of World War I veterans. He employed force and motion measurement techniques to help develop prosthetic limbs. Further research in this subcategory of kinesiology led to studying movement in three dimensions, which led to mathematical analysis of the range of joint forces, movements, and the elements of muscle strain and force against bone and ligaments.

The invention of the ergograph by Angelo Mosso in 1884 assisted greatly in further analysis of human movement in the first half of the twentieth century and evolved into several specialized forms to study very specific muscular functions in the human body. In general, the study of ergonomics as part of the Industrial Age push to better human work performance influenced the development of specialized sports equipment. Track shoes, for example, were designed to allow the foot a "natural" range of motion by using soft leather and a snug fit.

The study of aerodynamics also began to contribute to and employ the burgeoning knowledge of sports science. Likewise, biology has played a part as well: the study of marine mammals has helped shape the swimsuits and technique of many competitive swimmers.

More than any other technical innovation, however, the development of new materials used in sporting goods equipment changed the performance of both athlete and athletic gear. For the first half of the twentieth century, sports equipment was made primarily of steel, wood, or leather; but by the 1990s, composite materials such as graphite, fiber-glass, and new kinds of plastics, metals and ceramics were regularly employed. New metal composites, for example, dramatically altered the way bicycles operated. Virtually unchanged since the 1890's, the frame of bicycles has changed from the classic diamond design—the optimal shape for the strength capacity of steel, historically the most common material. With stronger, lighter materials such as titanium, high-tensile steel, chrome-moly steel, and aluminum, frame shapes could be altered to more aerodynamic shapes without losing strength.

Like other athletic equipment, shoe designs were limited mostly to sneakers, or rubber-soled canvas lace-ups, until the late 1960s. Designs changed and became more specific to sports—the high-top sneaker for basketball, for example—but little in the materials changed until the 1990s. Keeping in mind the lessons from biomechanics and kinesiology, shoe manufacturers tried to meet performance demand, protection, fit, and a saleable style. To protect the most vulnerable part of an athlete—the

musculoskeleton—the "impact waves" delivered by each foot strike to the ground need to be dampened. Shoe designers began to employ several different components in the outsole, insole, and midsole. A durable, flexible shock-absorbent rubber, such as elastomer styrene-butadeine is commonly used in outsoles. Insoles were typically made from polymer foam, and midsoles constructed of polymer foam called EVA, for ethyl, vinyl and acetate.

Running tracks also employed different forms of rubbers, polymers and plastics. The goal in track construction, as in shoes, is to reduce shock by increasing viscoelasticity. Other playing fields have evolved with new materials. High-strength tempered glass backboards have replaced most fiberglass, metal, and wood backboards in basketball. Protective gear in American football equipment involved several hard layers of plastic, wire, and soft foam. Astroturf, a plastic and rubber artificial turf, removed the threat of slippery, muddy fields but was not without controversy in some sports, particularly baseball, where "natural" grass was prized.

Like equipment technology, sports science today works largely on chemical innovation. Athletes monitor their nutritional input as much as they emphasize strength and endurance training. A low-fat, high-protein diet, for example, has been known to facilitate the repair of muscle tissue, and the intake of carbohydrates fuels the high-energy demands of a competitive athlete. Today the basics of biomechanics recorded in the early twentieth century have been applied to break down an athlete's movements into their component parts, pinpointing each muscle, nerve and tendon involved in given motion, which can then be targeted for improving performance. Strength training programs are designed for the muscle groups needed for a particular sport. Techniques for this vary, and continue to improve with the further understanding of physiology. For example, "plyometrics" is a popular strategy for pushing strength training beyond traditional weight-resistance techniques. Using a basic understanding of muscle contraction and response, plyometrics involves the stretching and contraction of muscles. Following a stretch, muscles contract more easily and responsively. A sit-up executed from an arched position over a ball, for example, will yield a stronger stomach than a standard sit-up. In the same vein, the importance of psychiatry and honing a "focus" preceding competition brought yoga into the regimen of many athletes.

In addition to increasing muscle strength, a further understanding of aerobic and anaerobic endurance has become a regular component of sports training. Even amateur athletes can now purchase products that measure their blood-oxygen level and provide information on the performance of their body. To push aerobic capacity, athletes alter their training to include intense, short episodes surpassing their peak aerobic mark. This technique—called anaerobic training—has been found to increase the body's aerobic capacity, thereby prolonging endurance at high levels of performance.

Today, the billion-dollar sports-equipment industry drives the majority of research and development in sports science. While the presence of performance enhancing drugs, such as steroids and testosterone derivatives, has been a legal bane, it has also shaped the way sports medicine and science perceive the function and capability of the human body in athletic events. It is likely that the use of drugs will continue to grow—illicit or not—alongside more accepted strategies.

See also **Health**

LOLLY MERRELL

Further Reading

Cicciarella, C. *Microelectronics in the Sport Sciences*. Human Kinetics, Champaign, IL, 1986.

Cavanaugh, P.R. *The Mechanics of Distance Running: A Historical Perspective*. Human Kinetics, Champaign, IL, 1990.

Fung, Y.C. Biomechanics: its scope, history, and some problems of continuum mechanics in physiology. *App. Mech. Rev*., 1–20, 1968.

Haag, H. *Sport Science in Germany: an Interdisciplinary Anthology*. Springer Verlag, Berlin, 1992.

Hoberman, J. *Mortal Engines: The Science of Performance and the Dehumanization of Sport*. Free Press, 1992.

Huxley, A.F. *Reflections on Muscle*. Liverpool University Press, Liverpool, 1980.

Pollack, G.H. *Muscles and Molecules: Uncovering the Principals of Biological Motion*. Ebner, Seattle, 1990.

Wilkerson, J.D. Biomechanics, The history of exercise and sport science. *Hum. Kinet*., 321–353, 1997.

Sputniks

The Sputniks were three artificial earth satellites launched by the Soviet Union in 1957 and 1958. They inaugurated the space age, were a tribute to Soviet science and technology, and caused surprise and panic in the U.S. similar to that after the Japanese attack on Pearl Harbor in December 1941. Space was transformed from a domain of fiction and fantasy into the "new frontier," ripe for scientific exploration, commercial exploitation, military penetration, and international rivalry.

The first Sputnik (the word is Russian for satellite) was launched on 4 October 1957 into a low-earth orbit. It was a small aluminum sphere 58 centimeters in diameter and weighed just under 84 kilograms. Sputnik-I circled the earth about every 96 minutes, burning out after 21 days as it plunged back into the atmosphere. Light reflected from its body, and the fact that it came as close as 228 kilometers to the earth, ensured that it was often visible to the naked eye. The frequency on which it emitted its "bip-bip" signal was easily captured by amateur radios. This public display of scientific and technological prowess was infused with political and ideological meaning. It was taken to demonstrate the superiority of Soviet communism over the capitalism of the West, and the U.S. in particular, and was intended to woo countries newly liberated from the yoke of colonialism into the Soviet camp.

Sputnik-II amplified the message. It was launched into an elliptical orbit on 3 November 1957 to commemorate the fortieth anniversary of the Bolshevik revolution. Cone-shaped, and 1.2 meters high, the satellite weighed 500 kilograms, carried scientific instruments and a small dog named Laïka. Laïka was wired for biomedical research, and she perished in her pressurized cabin after a few days in space. The satellite re-entered the earth's atmosphere on 14 April 1958.

The third and last Sputnik was launched on 15 May 1958. It weighed 13000 kilograms and carried a battery of scientific equipment. By now such exploits caused little stir. The U.S. had launched its first satellite on 31 January 1958 and another, with a different rocket, six weeks later. Even though rockets continued to pose enormous technical difficulties (no less than eight out of twelve U.S. attempts to launch satellites in 1958 ended in failure), it was evident that it was only a matter of time before satellite launches would become routine.

The first Soviet and American launches of scientific satellites into space were officially planned as part of their contributions to the International Geophysical Year (IGY). The IGY was a collaborative program in which scientists from 67 countries took advantage of a period of intense solar activity in 1957–1958 to study together a number of oceanographic and atmospheric phenomena. On 28 July 1955, the White House Press Secretary announced that the launch of small earth-circling satellites would be part of America's contribution to the IGY, providing "scientists of all nations [with an] important and unique opportunity for the advancement of science." Three days later the Soviets indicated that they had similar plans. Behind the U.S. offer was the intention to instrumentalize science for security. The Eisenhower administration wanted to begin a reconnaissance satellite program to spy on Soviet installations as part of the verification process of arms controls agreements. A civilian scientific satellite overflying Soviet territory would, it was thought, create a legal precedent for the "freedom of space," so opening the way for subsequent military-related space activities.

Moscow did not launch Sputnik without warning. An indication that the launch was imminent came on 26 August 1957, when President Khrushchev announced that the Soviet Union had successfully fired an intercontinental ballistic missile (ICBM) with a range of 8000 kilometers. Amateur radio magazines published the frequencies on which the satellite would emit its signal well before the launch date.

Sputnik shocked officials and the public in the U.S. because it was generally believed that the Soviet's were supplying misleading information for propaganda purposes. Once disabused of this misconception, the administration did not only have to deal with the blow to American prestige. It also had to face the fact, even more evident after Sputnik-II, that the Soviet rocket program had probably advanced to the point where ICBMs carrying nuclear weapons could strike New York from Moscow. Putting a dog in space also suggested that the Soviets were collecting data in anticipation of human space flight. The Sputniks were not simply scientific instruments and technological artifacts. They were also weapons in the Cold War rivalry between the superpowers, undermining American national pride, national security and presumed international technological supremacy.

The Sputniks were perceived differently in Western Europe, commensurate with local capabilities and the concerns of medium-sized world powers, and the different significance that space held on that side of the Atlantic. Scientists emphasized the scientific and technological achievements, and capitalized, when and where they could, on the opportunity space research offered for additional funding. The general public was fascinated by the sight and sound of Sputnik-I and, especially in Britain, deeply concerned about the fate of Laïka. (The Royal Society for the Protection of Cruelty to Animals' switchboard was jammed by indignant callers.) Governments, for their part, were relatively unmoved. Only a few, notably Britain and France, had small guided missile or

rocket programs, and these were restricted to the upper atmosphere; none had space ambitions. Sheltering under the U.S. nuclear umbrella, and sidelined by superpower rivalry, European governments needed other stimulants to propel them into space. It was the commercial possibilities, notably of telecommunications satellites, that eventually convinced most of them that a major space effort was needed. That was only clear a decade after the Sputniks first circled the globe.

See also **Rocket Propulsion, Liquid Propellant; Rocket Propulsion, Solid Propellant; Space Exploration, Unmanned**

JOHN KRIGE

Further Reading

Divine, R. *The Sputnik Challenge: Eisenhower's Response to the Soviet Satellites*. Oxford University Press, New York, 1993.
Killian, J. *Scientists, Sputniks and Eisenhower*. MIT Press, Cambridge, MA, 1977.
Le Grand Atlas de l'Espace. Encyclopedia Universalis, France, 1989.
Launius, R., Logsdon, J.M. and Smith, R.W., Eds. *Reconsidering Sputnik. Forty Years Since the Soviet Satellite*. Harwood Academic, Amsterdam, 2000.
McDougall, W. *The Heavens and the Earth. A Political History of the Space Age*. Johns Hopkins University Press, Baltimore, 1997.
Siddiqi, A. *Challenge to Apollo. The Soviet Union and the Space Age, 1945–1974*. NASA SP-2000–4408, Washington D.C., 2000.

Strobe Flashes

Scarcely a dozen years after photography was announced to the world in 1839, William Henry Fox Talbot produced the first known flash photograph. Talbot, the new art's co-inventor, fastened a printed paper onto a disk, set it spinning as fast as possible, and then discharged a spark to expose a glass plate negative. The words on the paper could be read on the photograph. Talbot believed that the potential for combining electric sparks and photography was unlimited. In 1852, he pronounced, "It is in our power to obtain the pictures of all moving objects, no matter in how rapid motion they may be, provided we have the means of sufficiently illuminating them with a sudden electric flash."

The electronic stroboscope fulfills Talbot's prediction. It is a repeating, short-duration light source used primarily for visual observation and photography of high-speed phenomena. The intensity of the light emitted from strobes also makes them useful as signal lights on communication towers, airport runways, emergency vehicles, and more. Though "stroboscope" actually refers to a repeating flash and "electronic flash" denotes a single burst, both types are commonly called "strobes."

The stroboscope consists of three basic components. First, a power supply—an electrical outlet or battery—sends electricity into a capacitor that collects the energy. When the stored energy is dumped into the third component, a tube containing electrodes and filled with a rare gas, the excited gas molecules produce a blast of brilliant light and heat. By adjusting the flashing rate, a moving object can appear to be:

1. Frozen in place, if the flash rate equals the speed of the object
2. Moving backward, if the flash rate exceeds the object's speed
3. Moving forward, if the flash rate is less than the speed of the object

The strobe's components originated in the work of eighteenth and nineteenth century experimenters who delved into the secrets of creating electricity and storing the resulting charge. Alessandro Volta built the first battery in 1800 from a stack of silver and zinc disks, each separated by a moist cloth. Another common generator, the Wimshurst machine, built up an electrostatic charge between two hand-cranked glass disks that rotated in opposite directions, but did not touch each other. The charge was collected and transferred by metal combs or chains to a Leyden jar, a glass jar lined with metal foil inside and out. Touching the inner and outer foil layers of this eighteenth-century capacitor simultaneously produce a powerful spark.

Michael Faraday pioneered the development of the final component of a stroboscope, the gas discharge tube, in 1839, using a glass jar and two electrodes (dubbed the "electric egg"). However, it was Johann Wilhelm Geissler who turned Faraday's experimental apparatus into a practical research tool in the 1850s by producing sealed, evacuated tubes containing platinum wires that served as electrodes.

By the early twentieth century, work on stroboscopes was reaching a practical level as generators, capacitors, and gas discharge tubes became more powerful, reliable, and easier to mass produce. A number of inventors contributed to the development of the strobe, including Etienne Oehmichen who, with the *Société Anonyme des Automobiles et Cycles Peugeot*, received French and Swiss patents around 1920 for a strobe-like device that he

specified would be used to examine motors in motion. In 1926 and 1927, Laurent and Augustin Seguin were awarded patents for their "flash-producing apparatus" by France, Switzerland, and the U.S. The brothers marketed this early strobe as the "Stroborama" and, by the early 1930s, the Seguins claimed industrial, government, and university laboratories around the world as their customers.

What made the stroboscope so useful was that the flash was renewable—as soon as the capacitor recharged, it was ready to go again; and since the flashing rate could be adjusted, it was possible to measure the speed of machinery in motion, as well as to spot any irregularities in the mechanism's operation, just by watching the machine in the rapidly-flashing light. However, devices like the Stroborama produced weak, long flashes that were not suitable for photography. Instead, it was the engineering and entrepreneurial talents of Harold Edgerton at the Massachusetts Institute of Technology (MIT) that made the strobe a commonplace device. Edgerton built on the existing work on strobes and experimented with ways to increase the flashing rate, make the flashes both brighter and shorter in duration, and produce light of the right color for photography.

Edgerton came to MIT as a graduate student in electrical engineering in 1926 to study the large motors used in power-generating plants. He understood the limitations of the neon strobes already in use, so he adapted a commercially available mercury tube, synchronized its short flashes to a motor's speed to make the spinning parts appear stationary, and shot his first stroboscopic photographs and motion pictures. Each brilliant flash lasted about 10 microseconds (1 microsecond $= 1 \times 10^{-6}$ seconds). From these first experiments, Edgerton created a variety of flash lamps and stroboscopes.

Throughout the late 1930s and 1940s, Edgerton concentrated on adapting the strobe for new applications. His images of stage shows and sporting events, illuminated by large strobes hung in theater and arena rafters, captured the public's imagination. These enormous flashes were later adapted for aerial nighttime reconnaissance photography during World War II. Edgerton's first camera-mounted, portable strobe ushered in a new era in photojournalism. Restless children became easier to photograph with the introduction of a strobe for studio photographers. To make high-speed motion pictures, Edgerton designed a camera without a shutter that rushed the film through continuously at speeds of around 23 meters per second. In a dark room, the regular flashing of the stroboscope acted like a shutter. Edgerton also created multiple-exposure still images of the essence of movement, such as the beat of a bird's wing. Working with explorer Jacques Cousteau, Edgerton developed strobes for underwater photography.

By the late 1940s, the demand for strobes supported more than 35 manufacturers in the U.S. During the 1950s, much of the design and manufacturing work moved first to Europe and then to Japan. Continual technical development has yielded shorter charging times, lighter-weight units, more automated functions, increased flash output, and improved circuitry, achieved by incorporating new semiconductors like transistors and later integrated circuits. The strobe has indeed become a standard tool in the photographer's kit.

See also **Cameras; Color Photography**

JOYCE E. BEDI

Further Reading

Bron, P. and Condax, P.L. *The Photographic Flash: A Concise Illustrated History*. Bron Elektronic, Allschwil, 1998.

Bruce, R.R., Ed. *Seeing the Unseen: Dr. Harold E. Edgerton and the Wonders of Strobe Alley*. George Eastman House, Rochester, NY, 1994.

Edgerton, H.E. *Electronic Flash, Strobe*. MIT Press, Cambridge, MA, 1979.

Edgerton, H.E. *Flash! Seeing the Unseen by Ultra High-Speed Photography*. Hale, Cushman & Flint, Boston, 1939.

Gus Kayafas, Ed. *Stopping Time: The Photographs of Harold Edgerton*. H.N. Abrams, New York, NY, 1987.

Talbot, W.H.F On the production of instantaneous photographic images. *Philosoph. Mag.*, 3, 15, 73–77, 1852.

Submarines, Military

Although the first military submarine operated as early as 1775 and development continued throughout the 1800s, the submarine was really a creature of the twentieth century; it was the submarine and the aircraft carrier that defined naval warfare in that century.

The basic technology of the submarine is quite simple and has remained constant since its inception. The boat submerges by taking on water through vents to decrease its buoyancy and surfaces by expelling the water with compressed air. The outward appearance of the military submarine has remained remarkably constant throughout its modern development—a cigar-shaped hull topped by

the immediately recognizable conning tower with a periscope for viewing the surface.

We can break submarine technology into five categories:

- Propulsion
- Hull design
- Weaponry
- Stealthiness
- Ancillary technologies.

The method of propulsion for the first half of the century was the diesel-electric system. Standard diesel engines were used for general operation on the surface but could not be used while submerged because of the enormous amount of air required for combustion. The submarine would only dive to attack or avoid detection, at which time the boat switched to power provided by electric batteries, charged from the diesels while on the surface. Most submarines were double hulled, with water filling the space between the two hulls while submerged.

The weapon that made the submarine useful was the self-propelled torpedo, powered by compressed air, which provided the sub with a deadly and stealthy attack. These technologies have been supported by countless others, each complex in its own right: atmosphere regeneration, escape and rescue techniques, underwater communications, weapons guidance, electronic countermeasure, and countless others, but all technologies unique to the military submarine revolved around one concept: avoiding detection by stealth.

The first significant use of submarines came in World War I when German U-boats attacked Allied shipping in the Atlantic Ocean. Losses of merchant ships increased and threatened to cripple the Allied war effort until a simple and ancient remedy was rediscovered—sailing the merchant ships in convoys. U-boats sank over 11 million tons of shipping but ceased to be a serious threat after the adoption of the convoy system.

In World War II, the critical Battle of the North Atlantic was a struggle defined by the technological accomplishments on each side as Germany's U-boat force sought to avoid detection from the eyes and ears of Great Britain's anti-submarine warfare (ASW) force. In the efforts at stealth and avoidance of detection, the most significant developments were the deployment of ASDIC (for Anti-Submarine Detection Investigation Committee), or sonar, followed closely by airborne radar. The most deadly enemy of the submarine turned out to be aircraft, which could detect surfaced submarines by means of radar allowing an attack with bombs or, as the submarine dived, depth charges. The U-boats countered British radar with a radar detector called Metox that warned of attack, but the British eventually deployed a new radar using a centimetric wavelength undetectable by Metox. The Germans did not discover the use of the new radar and were slow to develop an effective counter.

U-boat losses continued to rise. As a stopgap measure the Germans deployed the schnorkel, developed before the war by the Dutch but captured by the Germans upon the surrender of the Netherlands. The schnorkel, or snorkel to the Americans and snort to the British, was a simple device—a breathing tube that could be raised similar to a periscope that allowed the submarine to run its diesel engines while submerged. While technologically interesting, its practical deployment was a failure; Allied radar could soon detect even the schnorkel protruding from the water.

The war ended before the Germans could deploy their own next wave of technology embodied by the Type XXI "Walther" boat, powered by hydrogen peroxide fuel and much larger battery capacity that gave it a fast underwater speed.

A submarine is only as effective as its weaponry is reliable, and World War II saw examples of massive weapons systems failures. In the Norwegian campaign of 1940, the earth's magnetic field interfered with the operation of the U-boats' magnetically armed torpedoes. German U-boats operating off the Norwegian coast aimed torpedoes at unsuspecting British capital ships only to hear their duds clank off the sides of the targets.

In the Pacific, American submarines were armed with hopelessly defective torpedoes that rendered the American submarine fleet useless for many months until the flawed torpedo design was corrected. When effective torpedoes reached the American subs, their effect was devastating. The Japanese never developed an effective ASW force or doctrine, and U.S. subs ran wild, destroying over 60 percent of Japanese merchant shipping and paralyzing the import-constrained Japanese economy. While less well known than the great carrier battles and island invasions, the U.S. submarine force contributed at least as much to the defeat of Japan.

The most significant single development in submarine technology has undoubtedly been the use of nuclear propulsion. The first nuclear-powered boat was the *USS Nautilus*, launched in 1955. Nuclear power freed submarines from the need to surface or schnorkel. Subs could stay at sea for months longer than before and stay submerged indefinitely. The drawback of nuclear power was that it was relatively noisy, and a new generation of

diesel submarines remained in use through the remainder of the century, particularly in short-range roles. The submarine also entered the area of strategic nuclear warfare, as it provided an ideal platform for long-range missiles tipped with nuclear warheads. The first launch of a ballistic missile from a submarine came from the *USS George Washington* in 1960.

With the security of a nation's entire population dependent on its defense against enemy ballistic missile submarines, antisubmarine and stealth technology became even more important. Both the U.S. and the Soviet Union devoted enormous amounts of resources in research in the race to detect the other's subs and protect their own. If the ASW-submarine contest of World War II was a battle of technology, the competition between Cold War fleets was even more so. Hull design improved tremendously; World War II subs could dive to 120 meters, and modern submarines can reach much greater depths. Hulls were also more streamlined, further increasing speed.

Weaponry diversified from earlier years. In addition to ballistic missiles and the traditional torpedo, subs began deploying sophisticated cruise missiles, first for attacking surface naval targets and, with the introduction of the American Tomahawk cruise missile, land targets.

The airplane, while still useful, gave way to the submarine itself as the most effective antisubmarine weapon with hunter-killer submarines on both sides patiently stalking the other's missile submarines lurking deep in the ocean as far as possible from enemy bases. The use of active emission sonar fell from favor except for targeting immediately before an attack, as its use gave away the position of the attacking sub. Passive listening sonar became the preferred method of tracking an enemy boat, and hence silence became the most important defense for the submarine.

A representative example of the many developments was when ASW forces began finding subs by the use of magnetic anomaly detection (MAD), the Soviet constructed their Alfa class with hulls of nonmagnetic titanium at great expense. Unfortunately for the Soviets, the machinery of the Alfas was so noisy they could be easily located by passive sonar. Such tradeoffs and competitions existed in all facets of submarine and ASW technology.

With the end of the Cold War, the usefulness and cost-effectiveness of the nuclear submarine began to be seriously questioned, but technological advances continued. The *USS Seawolf*, launched in 1997, boasted a nuclear reactor fueled with liquid sodium and the Virginia class promised to be even more advanced when deployed early in the twenty-first century.

The use of the submarine and its technological advance paralleled the changes of war in world society over the course of the century. World War I began with the U-boats operated by a strict law of maritime warfare, warning merchant ship crews of their presence and patiently waiting for the embarkation of the crew in lifeboats. Well before the end of the century, nuclear submarines lurked in the depths, each waiting to destroy dozens of enemy cities without warning.

See also **Radar Aboard Aircraft; Sonar; Submersibles**

FRANK WATSON

Further Reading

Clancy, T. *Submarine: A Guided Tour Inside a Nuclear Warship*. G.P. Putnam's Sons, New York, 1993.
Davies, R. *Nautilus: The Story of Man under the Sea*. Naval Institute Press, Annapolis, MD, 1995.
Miller, D. and Jordan, J. *Modern Submarine Warfare*. Military Press, New York, 1987.
Van der Vat, D. *Stealth at Sea: The Submarine at War*. Weidenfeld & Nicolson, London, 1994.

Submersibles

Beginning in the 1860s, and up to World War I, most submersible inventions were marketed for their military potential. After World War I, however, interests in oceanic exploration, for purposes ranging from scientific research to resource exploitation and wreckage recovery, prompted the development of several types of civilian submersibles. A further reason for the development of submersibles was an understanding of the limit of the human body to withstand great depths. Initially, diving bells were the basic means of underwater exploration, but by the twentieth century it was clear that no amount of improvement to the open diving bell (where the air became more compressed as the depth increased) could compete with enclosed habitats.

Enclosed bells were bathyspheres (from the Greek "deep sphere"), submersibles tethered to the surface for suspension, air, and power supply, first developed in the late nineteenth century. In 1930, naturalist William Beebe and geologist Otis Barton designed a stainless steel bathysphere that was 1.45 meters in diameter, with walls 3.8 centimeters thick, and three windows of quartz glass for observation purposes. The inside equip-

ment was spartan, with no seat for the two occupants, internal oxygen tanks with 8 hours of supply, and soda lime and calcium chloride chemicals spread to absorb carbon dioxide (CO_2) and moisture. First diving to 244 meters in 1930, the two scientists eventually established a record of 923 meters in 1934. This record would hold until after World War II, partly because of limited interest in and therefore limited funding for oceanographic science.

In the meantime, a Swiss physicist, Auguste Piccard, had successfully navigated to 18 kilometers into the stratosphere in a sealed sphere, FNRS-1, attached to a balloon (the Belgian National Science Foundation had paid for the craft, hence the French language acronym that christened it). Based on his flight experience, Piccard conceived of a similar plan, this time to reach great oceanic depths. He felt, however, that the bathysphere was too vulnerable due to its dependency on the tether, which might snap. Thus, Piccard and his son, Jacques, decided to include a float that would have a function similar to that of a balloon. The float featured a gasoline tank (gasoline is incompressible and lighter than water, which would help raise the vehicle to the surface). Several tons of iron pellets were also in the float to provide negative buoyancy. Although the craft, known as the bathyscaphe ("deep boat") FNRS-2, was eventually launched to a depth of 4048 meters in 1954. Piccard was no longer involved in the project, after a falling out with the French Navy, which had taken over the project from Belgium. He therefore shifted his interest to a new project, a bona fide exploration submarine.

This new model, the Trieste (named in honor of the Italian city that had funded its construction), was divided into two sections, with the upper containing 106 cubic meters of gasoline, two ballast water tanks, and 9.1 tons of metal pellets (later 16 tons). The second section was a steel alloy sphere large enough for two people. The whole contraption weighed 50 tons. Although built for Auguste Piccard and his son Jean and modified in the 1950s, it was eventually transferred to the U.S. Navy, which helped Jean Piccard and US Navy Lt. Don Walsh set a depth exploration record in January 1960 by reaching a depth of 10.91 kilometers in the Mariana trench. Trieste's exploit, as well as its assistance in discovering the wreckage of the *USS Thresher* submarine off Massachusetts in 1963 reflected a shift in military and governmental circles in favor of oceanographic research and recovery. Allyn Vine of the Woods Hole Oceanographic Institute had recommended the purchase of Trieste by the U.S.. His contribution to the development of research submersibles would soon be acknowledged through the naming of a new type of deep research vessel (DRV) in his honor.

The Alvin (named for Allyn Vine) is one of the first and longest serving DRVs. Operated by the Woods Hole Oceanographic Institute since 1965 and weighing 17 tons, it uses six reversible electric thrusters, can carry a payload of 680 kilograms at a maximum speed of 2 knots. Its crew of three can remain submerged for up to 72 hours in the titanium pressure hull. Like other research submersibles of its kind, Alvin requires a support ship to remain in operation, as its maximum cruising range is limited to five kilometers. Alvin's early successes (it helped locate a lost H-bomb off the Palomares coast of Spain) also prompted a "golden age" of civilian submersible design in the 1960s, when several commercial firms took to designing DRVs. Westinghouse built the Deepstar 4000 for oceanographer Jacques Cousteau, while Grumman designed the PX-15 Ben Franklin, a mesoscaphe for use in exploring mid-depth phenomena such as the Gulf Stream. Although mostly of steel and titanium, some crafts came to include acrylic plastics that allowed crew to have a greater view outside the observation compartment.

The first submersible to be used for tourism was the Auguste Piccard, capable of carrying 40 passengers and used during the Swiss National Exhibition in Lausanne in 1964 to make dives in Lake Geneva. Later models, much smaller, are still used for underwater viewing. Canadian and French companies did attempt the construction of a long-range commercial nuclear-powered submarine, SAGA-1, capable of carrying 15 passengers for a month to investigate oil drilling and other oceanic resource exploitation. However, financial problems suspended the project in 1987.

DRVs also rely on the use of remotely piloted vehicles (RPVs). Although limited in their steering and recovery capacities, these machines began playing an important role in tracking underwater wrecks. RPVs usually feature a steel frame that includes the necessary instruments and an engine, but more advanced types feature tracks for crawling on the sea bottom. The most famous were the ones used by Dr. Robert Ballard to locate the wreck of the *Titanic* in 1985–1986, which included the Jason, piloted from Alvin.

Small civilian submarines are commonly used in industrial surveys and underwater tourism, but the infrastructure required for bigger ships makes their

cost prohibitive to the extent that governments are either the main sponsors or even the main operators of these ships.

See also **Submarines, Military**

GUILLAUME DE SYON

Further Reading

Ballard, R.D. *The Discovery of the Titanic*. Warner/Madison Press, New York, 1987.

Beebe, W. *Half Mile Down*. Harcourt Brace, New York, 1934.

Geyer, R.A. *Submersibles and Their Use in Oceanography and Ocean Engineering*. Elsevier, New York, 1977.

Piccard, J *The Sun Beneath the Sea*. Charles Scribner's Sons, New York, 1971.

Shenton, E.H. *Exploring the Ocean Depths*. W.W. Norton, New York, 1968.

Superconductivity, Applications

The 1986 Applied Superconductivity Conference proclaimed, "Applied superconductivity has come of age." The claim reflected only 25 years of development, but was justifiable due to significant worldwide interest and investment. For example, the 1976 annual budget for superconducting systems exceeded $30 million in the U.S., with similar efforts in Europe and Japan. By 1986 the technology had matured impressively into applications for the energy industry, the military, transportation, high-energy physics, electronics, and medicine. The announcement of high-temperature superconductivity just two months later brought about a new round of dramatic developments.

By 1986 the energy industry witnessed development of large superconducting projects for fusion power generation, magnetic energy storage, transmission lines, and industrial motors and generators. For instance, the international large coil test, begun in 1977 to demonstrate the capability of producing fields sufficiently strong for fusion power, utilized six D-shaped superconducting magnets, each 4.5 meters tall, 3.5 meters wide, and 1 meter thick. The coils, developed by Japan, Switzerland, Euratom (a European consortium) and three U.S. industries, were toroidally assembled at Oak Ridge National Laboratory where they successfully produced a 9-tesla field in 1987. Enormous superconducting solenoids were designed in the U.S. and Japan to store energy for diurnal load leveling by electric utilities. The designs for 5000 megawatt-hours storage proposed magnets 1 kilometer in diameter by 20 meter tall, supported by an underground earth-based structure. Superconducting transmission lines such as the 1000 MVA prototype developed from 1971–1986 at Brookhaven National Lab were designed for underground service to mitigate power congestion in large cities.

The British and U.S. Navies utilized superconducting magnets to build homopolar motors for ship propulsion. Although only 2 horsepower in 1965, these motors grew to 3250 horsepower by 1970 in the U.K. and to similar size by 1983 in the U.S. High-speed trains levitated with superconducting magnets exceeded 500 kilometers per hour by 1980 on the Japanese Miyazaki test track. In the field of high-energy physics, superconducting dipole magnets were built to accelerate beams of subatomic particles around multi-kilometer long rings, while large superconducting solenoid magnets were constructed to detect the particle spray from collisions of the accelerated beams. The "energy-doubler" accelerator built at Fermilab in Batavia, Illinois between 1971 and 1983 utilized 744 6-meter-long dipoles, and 244 2-meter-long quadrapole magnets.

Magnetic resonance imaging (MRI) provided the largest commercial success for superconductivity. Introduced in 1980, by 1986 MRI systems were used in over 700 hospitals and represented a $1 billion per year industry. The very stable 1–2 tesla field produced by the superconducting MRI solenoids coupled with sophisticated computer systems enabled a noninvasive technique for seeing inside the body (see entry on Nuclear Magnetic Resonance). The second major medical application of superconductivity utilized their ability to detect magnetic fields as small as 10^{-14} tesla. Medical devices using superconducting quantum interference devices (SQUIDs) were developed during the 1980s to detect the fields arising from miniscule currents within the human brain and heart, allowing nonsurgical diagnoses of irregularities. Josephson junctions, the building block for SQUIDS, were also utilized to develop high-precision high-speed electronics. Projects in the U.S. and Japan explored the possibility of superfast computers, developing logic circuits by the mid-1980s with switching times of only 2.5 picoseconds.

Numerous technical challenges were overcome to achieve the remarkable progress reported above. Constructing reliable superconducting magnets required breakthroughs in five different areas: structural support, conductor stability, protection systems, AC losses, and optimized material microstructures. Structures were developed to support pressures exceeding 100 atmospheres produced inside the coils from the combination of high

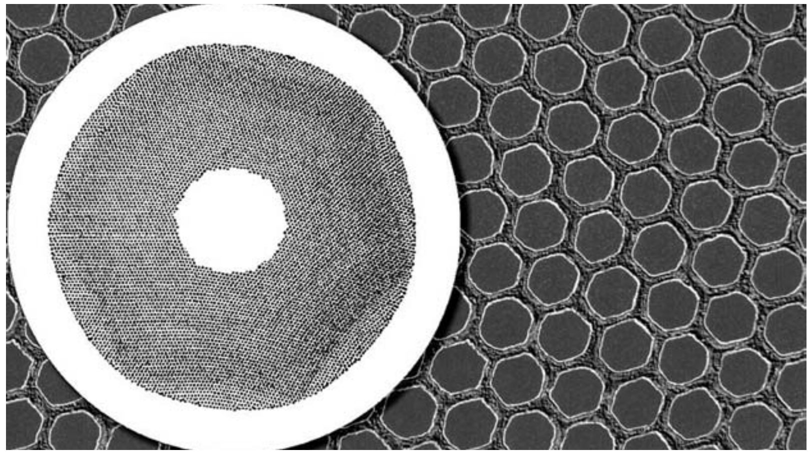


Figure 29. Enlargement of 0.8-millimeter diameter NbTi composite superconductor fabricated in 1992 by the Advanced Superconductor division of Intermagnetics General Corporation. The cross section of the conductor is superimposed on an enlarged detail of the 6-micrometer diameter filaments. The NbTi filaments are surrounded by a high-purity copper matrix. [*Source: Image by Peter J. Lee. Used with permission*.]

currents and high fields. Conductor design principles were established so that thermal energy, released when large forces caused conductor motion, would not cause permanent loss of superconductivity. Magnet protection systems were devised to avoid permanent damage to the coils if large portions became nonsuperconducting. Conductor cables were twisted to minimize losses associated with changing magnetic fields. Multistage conductor fabrication processes involving bundling, swaging, and heat treatments were developed to maximize the current-carrying capacities. Maximum current densities for niobium–titanium (NbTi) conductors rose from 2000 amps per square millimeter at 4.2°Kelvin and 5 tesla in 1980 to 5000 amps per square millimeter by 1991. As shown in Figure 29, an optimized NbTi conductor was made up of thousands of 10-micrometer-sized filaments embedded in copper.

The discovery of high-temperature superconductivity (HTS) in 1986–1987 presented the exciting possibility of simplified cooling requirements, but introduced an entirely new set of technical challenges. Foremost among these was the materials problem of converting ceramic superconductors into practical wires, tapes, or films. Although large superconducting currents were measured within the plane of single grains, by 1991 researchers identified that intergranular current flow was very sensitive to grain alignment and grain-boundary cleanliness. Successful processing techniques required very pure original ingredients, and the means to produce flattened gains with large intergranular contact surfaces. Among the many HTS materials, bismuth–strontium–calcium–copper–oxide BSCCO) and yttrium–barium–copper–oxide (YBCO) have demonstrated the most promising properties for applications. The powder-in-tube approach, introduced by the Japanese in 1989 demonstrated a reliable method to produce silver-clad BSCCO conductors. Here, superconductor-precursor powders are packed into silver tubes and subsequently subjected to a sequence of heat treatments and deformation steps. Multifilamentary versions of these conductors, first developed by Sumitomo Electric in 1991 are now widely used in emerging applications. In 1995 Los Alamos National Lab reported critical current densities exceeding 1 million amps per square centimeter at 75°Kelvin in thin-film tapes of YBCO. Because it can support higher current densities than any other superconductor at 75°Kelvin and at fields larger than 20 tesla, considerable effort is underway to produce useful conductor lengths with this material.

Since 1987 worldwide development of HTS applications has been increasing rapidly. Examples of 5000-horsepower motors, 1.2 MVA transmission lines, and MRI coils mirror the low-temperature superconducting (LTS) applications. Although the number of large-scale LTS projects has been dwindling since 1986, new HTS applications have emerged. The levitating strength of HTS materials is being utilized in flywheel energy storage systems, telescope tracking structures, and maglev transportation vehicles. Combined with compact refrigerators, HTS filters are being utilized in cellular-phone base stations and satellite communication systems.

See also **Josephson Junction Devices**

JOHN PFOTENHAUER

Further Reading

Birmingham, B.W. and Smith, C.N. A survey of large scale applications of superconductivity in the US. *Cryogenics*, 16, 59–71, 1976.

Dahl, P.F. *Superconductivity: Its Historical Roots and Development from Mercury to the Ceramic Oxides*. American Institute of Physics, New York, 1992.

Hulm, J.K. International Cooperative—Collaborative Perspectives: Superconductive Science and Technology. *IEEE Trans. Magn.*, MAG-23, 2, 423–426, 1987.

Sheahen, T.P. *Introduction to High Temperature Superconductivity*. Plenum Press, New York, 1994.

Simon, R. and Smith, A. *Superconductors: Conquering Technology's New Frontier*. Plenum Press, New York, 1988.

Weinstock, H. and Nisenoff, M., Eds. *Microwave Superconductivity*. Kluwer, Dordrecht, 2001.

Wilson, M.N. *Superconducting Magnets*. Oxford University Press, Oxford, 1983.

Superconductivity, Discovery

As the twenty-first century began, an array of superconducting applications in high-speed electronics, medical imaging, levitated transportation, and electric power systems are either having, or will soon have, an impact on the daily life of millions. Surprisingly, at the beginning of the twentieth century, the discovery of superconductivity was completely unanticipated and unimagined.

In 1911, three years after liquefying helium, H. Kammerlingh Onnes of the University of Leiden discovered superconductivity while investigating the temperature-dependent resistance of metals below 4.2°Kelvin. Later reporting on experiments conducted in 1911, he described the disappearance of the resistance of mercury, stating, "Within some hundredths of a degree came a sudden fall, not foreseen [by existing theories of resistance]. Mercury has passed into a new state, which ... may be called the superconductive state." By February of 1914 Onnes discovered that tin and lead were also superconductors, and that all three elements remained superconducting only below their threshold temperature, current, and magnetic field. In 1913 Onnes proposed production of intense magnetic fields, 200,000 times larger than the earth's magnetic field of 0.5 gauss, by winding a 30 centimeter diameter coil out of superconducting lead wire. Unfortunately the idea was not to be realized at that time. As the current, and resultant magnetic field were increased, the wire became resistive at a much lower current than the threshold current in zero magnetic field. In 1917 Silsbee explained the reduced performance by relating the maximum allowable current to the presence of the superconductor's threshold magnetic field at the conductor surface. Of the many superconductors known today, the 24 that are pure elements all lose their superconductivity at magnetic fields of less than 0.2 tesla (2000 gauss); far below the multitesla fields proposed by Onnes.

Another surprise in the field of superconductivity came in 1933 when Meissner and Ochsenfeld found superconductors to behave differently than perfect conductors with respect to their interactions with magnetic fields. As shown in Figure 30, when a perfect conductor (here, a solid sphere) is cooled in the presence of a magnetic field, and the field is subsequently removed, surface currents are established to maintain the magnetic field in the sphere. However, cooling a superconductor below its transition temperature in the presence of a magnetic field, causes the field to be immediately expelled from the superconductor. Such behavior motivated Gorter and Casimir in 1934 to describe superconductivity as a separate thermodynamic state, in a similar sense that ice is a separate thermodynamic state of water.

The investigation of alloy superconductors between 1930 and 1935 presented further perplexities. For these materials, two critical magnetic fields H_{c1} and H_{c2} were identified. Below H_{c1} the alloys behaved as the elemental superconductors, expelling the magnetic field. However, between H_{c1} and H_{c2}, magnetic flux, and associated nonsuperconducting or "normal" regions, increasingly penetrated the superconductors until at H_{c2} superconductivity was completely eliminated. In 1950, Pippard at Cambridge explained why supercon-

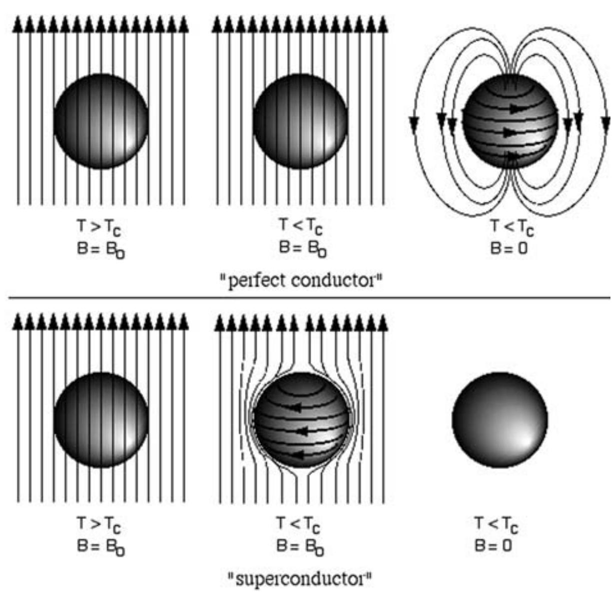


Figure 30. The response of a perfect conductor (above) and a superconductor (below) to the sequence: apply magnetic field, cool below the critical temperature, remove magnetic field.

ducting elements and alloys behave differently basing this on the surface energy between the superconducting and normal regions. For reasons similar to those dictating whether water will spread-out or bead-up on various surfaces, Pippard showed that the elements favored a minimum surface between the superconducting and normal regions, thereby expelling all magnetic flux, while the alloys favored maximum interfacial surface, thereby enabling flux penetration. In 1957 Abrikosov at Moscow accurately predicted that the penetration of magnetic flux would take the form of a regular array of flux lines. Also in 1957, a microscopic explanation for the existence of superconductivity was provided. The BCS theory, named for Bardeen, Cooper, and Schrieffer, of the University of Chicago, accounted for the absence of resistance by identifying a coupling of electrons in a low-energy state. In the coupled state, usually dissipative collisions with the vibrating atomic lattice became shared exchanges of momentum.

During the 1950s a foundation was quietly being laid for an explosion of applications in the 1960s. In 1955, Yntema at the University of Illinois utilized an advantage of current density in a strained sample of Niobium wire to wind a small magnet that produced an unexpectedly high field of 0.7 tesla. Hulm at Westinghouse, followed quickly by others, recognized the importance of the strained, or cold-worked, material. By 1961 magnets were developed using NbZr, NbTi, and Nb_3Sn that produced fields as high as 7 tesla. All were surprised that the conductors were able to carry such large currents. An understanding developed gradually through the 1960s that for the alloy, or type II, superconductors the maximum current density was not intrinsically linked to the magnetic field. Rather, the maximum current was dependent on how well the flux lines were pinned by microstructural imperfections in the material. Following this understanding, numerous magnets were developed that produced the fields envisioned by Onnes in 1913. Two developments in 1956 also birthed a variety of superconducting electronics applications. In the first, a layered structure of tin, insulator, and lead was utilized as a fast-switching element in a digital computer. The second, an observation by Josephson of superconductive "tunneling" through insulators, enabled alternative fast-switching devices, and superconducting quantum interference devices (SQUIDs) that allow extremely sensitive measurements of magnet fields.

A final explosion in superconductivity occurred in 1986. Convinced the search for superconductivity in the intermetallic compounds should not be further pursued, Bednorz and Müller at IBM in Zurich conducted a four-year search through a class of ceramic materials called perovskites. Their cautious announcement in September of 1986 entitled "Possible High T_c Superconductivity in the Ba–La–Cu–O System" described a precipitous drop in resistance at 35°Kelvin, more than 10°Kelvin higher than any previous superconductor. A flurry of reports during the next two years from the U.S., Japan, and China revealed additional high-T_c superconductors with critical temperatures ranging up to 150°Kelvin. The difficult task of converting the ceramic materials into reliable electronics devices and wires for magnets occupied engineers and scientists for the remainder of the twentieth century. Their determination is now providing an exciting array of commercial applications.

See also **Superconductivity, Applications**

JOHN PFOTENHAUER

Further Reading

Dahl, P.F. *Superconductivity: Its Historical Roots and Development from Mercury to the Ceramic Oxides*. American Institute of Physics, New York, 1992.
Hazen, R.M. *The Breakthrough: The Race for the Superconductor*. Simon & Schuster, New York, 1988.
Lynton, E.A. *Superconductivity*. Methuen, London, 1969.
Mendelssohn, K. *The Quest for Absolute Zero: the Meaning of Low Temperature Physics*. Taylor & Francis, London, 1977.
Schooley, J.F., Ed. Applied Superconductivity Conference, 1986. *IEEE Trans. Magnet*., MAG-23, 2, 1987. The first nine articles in this volume comprise a session celebrating the seventy-fifth anniversary of the discovery of superconductivity and recounting its history.
Sheahen, T.P. *Introduction to High Temperature Superconductivity*. Plenum Press, New York, 1994.
Simon, R. and Smith, A. *Superconductors: Conquering Technology's New Frontier*. Plenum Press, New York, 1988.

Surgery, Plastic and Reconstructive

As early as 1500 BC, surgeons in India used leaves to rebuild the amputated noses of criminals. By the Renaissance, Tagliacozzi devised a flap and graft using the patient's skin. A hiatus then prevented further development of plastic surgery technologies because of a belief that restoring and remolding was disreputable in that it interfered with the artistry and will of God. This attitude lasted until after World War I when plastic surgery came of age as an established specialty. The dire need for reconstruction and cosmesis from the mutilating defects of war spurred the research of new technologies.

In 1917, the tubed pedicle flap was devised simultaneously by Sir Harold Gillies (London) and W. Filatov (Ukraine). The graft could now be freed to grow independently. The dermatome, a surgical device that allowed the operator to excise tissue of uniform and predetermined thickness, was invented by Padgett in 1939 and used extensively in plastic surgery.

Wartime morbidity and industrial accidents challenged surgeons to restore circulation to damaged body parts. The first successful reattachment of an arm was reported in 1962 and of a finger in 1968. The 1980s saw the rise of microsurgery where small vessels, muscles, bone, nerves, and veins could be repaired because microscopy allowed adequate visualization for manipulation of tissues without damage to tiny structures. Repair of cleft lip and cleft palate and other congenital anomalies, including the genitalia, were improved with microsurgery. Chemical and mechanical technology contributed to improved healing through the development of synthetic antibiotics and safer and longer lasting anesthetics. Early antibiotic use hinged on Fleming's discovery of *Penicillium notatum* in London (1928), which was followed by tyrothricin (Rene DuBos) and Waksman's development of streptomycin in 1944.

After World War II, small-breasted Japanese women, emulating the images of American women, underwent breast augmentation surgery using silicone injection. Failure of the procedure resulted in granulomas and silicone migration. This led to the development of a gel prosthesis in 1963 by Cronin and Gerow with a Dacron patch for attachment to the chest wall. A variety of breast implants continued to be used, some with smooth outer envelopes, others with a fuzzy polyurethane shell believed to stimulate tissue retention. Internal materials ranged from dimethylsiloxane nonliquid gel to saline. One technology used tissue expanders for the gradual introduction of saline over a period of time. The DIEP procedure is a microsurgical approach to breast reconstruction using the patient's own skin and fat. After mastectomy, a flap is made from abdominal tissue, and after blood flow is established it is transplanted to the chest wall site.

Cosmetic facial surgery gained in acceptance for the general public as the media promoted stars who appeared to defy ageing. The rhinoplasty, originally developed in Europe, gained popularity as second and third generation Jewish and Italian women in the U.S., desirous of assimilation, underwent this surgery. Autogenous graft materials were taken from ear or rib cartilage, septal tissue, and rib or iliac bone. Gore-Tex, a synthetic material, has been used to augment facial soft tissue.

During the last quarter of the twentieth century, tissue expanders and endoscopy (1993) contributed to less operative time and trauma, faster healing, and reduced intraoperative bleeding. "Facelifts" consist of many procedures. Brow and forehead lifts remove wrinkles from the upper face. Endoscopy removes excess skin and fat from around the upper and lower eyelids. Implants of collagen are used to enhance cheekbones. In the 1990s came the startling news that *Clostridium botulinum*, the same organism known to cause botulism, was being used in cosmetic surgery. The organism produces two toxins, type A and type B. Tiny amounts of these as Botox are injected into selected small facial muscles to paralyze them so that frowning will not be possible. Some doctors use electromyograms to guide the needle in the procedure. A self-inflating expander that contains salt, which gradually absorbs fluid, was also being researched at the turn of the twenty-first century.

Perhaps the most ubiquitous technology has been the laser, beginning in 1961 for ophthalmology, then introduced to plastic surgery in the 1990s. The original ruby crystal was replaced with argon and then the pulsed-dye laser. In 1995, laser resurfacing using carbon dioxide (CO_2) was introduced. The tremendous use of lasers within the field of plastic surgery has resulted in removal of hemangiomas, tattoos, vascular lesions, and wrinkles. Endoscopic surgery combined with the erbium: yttrium–aluminum–garnet laser is another late-twentieth century technology.

Between 1930 and 1990, chemicals such as resorcinol, salicylic acid, phenol, and trichloracetic acid (TCA) peels were used to smooth skin. Dermabrasion, a mechanical process, was also employed. Dressings, adhesives and masks to aid healing developed with each of these technologies.

Liposuction has been used since the late 1970s. This technique uses a cannula and suction equipment to remove fat from the thorax, abdomen, and extremities. Internal ultrasound is combined with suctioning so the operator can better view tissue planes.

For most of the twentieth century, cutting tools were made from metal and ceramic materials. The latest development is the argon gas torch, which both creates an incision and coagulates the blood to limit bleeding. Technologies that have greatly enhanced the ability to stop capillary bleeding during cosmetic surgery are platelet gel and fibrin glue, which seal wound surfaces and stop bleeding

during surgery and the formation of postoperative hematomas. Superior to commercial products, the patient's own blood cells can be collected and used to preclude disease transmission and problems of tissue incompatibility. Suture materials developed from silk and gut have been replaced with vicryl and absorbable synthetic materials; needles are preloaded and packaged in sterile containers.

Body fashion, like other cultural phenomena, changes with time, place, and value. It appears that the more the human body is exposed to display, the greater a variety of technologies will develop to mold it into culturally pleasing icons.

LANA THOMPSON

Further Reading

Biggs, T.M. and Humphreys, D.H. Augmentation mammaplasty, in *Grabb and Smith's Plastic Surgery*, Smith, J.W. and Aston, S.J. Eds. Little, Brown & Company. Boston, 1991.

Castiglioni, A. *A History of Medicine*. Alfred A. Knopf, New York, 1958.

Haeger, K. *History of Surgery*. Bell, New York, 1988.

Jankaukas, S., Kelman, C.I. and Grabb, W.C. Basic technique of plastic, in *Grabb and Smith's Plastic Surgery*, Smith, J.W. and Aston, S.J., Eds. Little, Brown & Company. Boston, 1991.

Man, D., Plosker, H. and Winland-Brown, J. The use of autologous platelet-rich plasma (platelet gel) and autologous platelet-poor plasma (fibrin glue) in cosmetic surgery. *Plast. Reconstr. Surg*., 107, 229–239, 2001.

Margotta, R. *The History of Medicine*. Smithmark, 1996.

Roberts, T. and Pozner, J.N. Aesthetic laser surgery in *Plastic Surgery: Indications, Operations and Outcomes*, Guyuron, B., Ed. Mosby, St. Louis, 2000.

Synthetic Foods, Mycoprotein and Hydrogenated Fats

Food technologists developed synthetic foods to meet specific nutritional and cultural demands. Also referred to as artificial foods, synthetic foods are meat-free and are designed to provide essential fiber and nutrients such as proteins found in meats while having low saturated fat and lacking animal fat and cholesterol. These foodstuffs are manufactured completely from organic material. They have been manipulated to be tasty, nutritionally sound with major vitamins and minerals, have appealing textures, and safe for consumption. Synthetic foods offer people healthy dietary choices, variety, and convenience.

Mycoprotein

Mycoprotein is created from *Fusarium venenatum* (also known as *Fusarium graminearum*), a small edible fungi related to mushrooms and truffles that was initially found in the soil of a pasture outside Marlow in Buckinghamshire, England. Concerned about possible food shortages such as those experienced in World War II Europe; as global populations swelled postwar, scientists began investigating possible applications for this organism as a widely available, affordable protein source. Scientists at one of Britain's leading food manufacturers, Rank Hovis McDougall, focused on mycoprotein from 1964. At first, they were unable to cultivate fungus to produce mycoprotein in sufficient quantities for the envisioned scale of food production. Food technologists devoted several years to establishing procedures for growing desired amounts of mycoprotein. They chose a fermentation process involving microorganisms, somewhat like those historically used to create yogurt, wine, and beer.

A controlled fermentation process permits consistent production and harvesting of mycoprotein. Unlike alcohol fermentation, mycoprotein fermentation retains the microorganism cells instead of discarding them when the process is completed. The process begins with the sterilization of a fermenter container to protect microorganisms from harmful contaminants. Technologists provide the microorganism culture with glucose to sustain growth and expansion. Ideal fermenting conditions for mycoprotein production require sufficient oxygen and nitrogen, temperature monitoring, and adequate glucose, biotin, and mineral nutrient supply. Steady, sustained mixing ensures suitable oxygen and food levels. For small fermenters, magnetic stirrers or paddles are used to achieve mechanical stirring of the culture. Large-pressure cycle fermenters function by appropriating air movement inside the fermenter to mix microorganisms. When the desired amount of cells has been generated, technologists remove them from the fermenter. Harvested mycoprotein resembles soft bread dough and is bland tasting. Consisting of fine fibers, fermented mycoprotein has a texture reminiscent of lean meats. Approximately 50 percent of mycoprotein is protein.

British nutritionists selected mycoprotein as the main ingredient to create food that was eventually given the brand name Quorn™. For a decade, they evaluated mycoprotein's safety with such studies as feeding it to a succession of test subjects, beginning with rats and dogs, and finally humans. Several generations of animals that were fed mycoprotein diets flourished. Human digestibility trials were also successful, and the U.K. Ministry of Agriculture, Fisheries and Food designated mycoprotein safe for nutritional consumption. Beginning in 1985,

Marlow Foods Ltd, a subsidiary of the pharmaceutical manufacturer AstraZeneca, used mycoprotein to manufacture a variety of Quorn foodstuffs flavored to substitute for chicken and beef.

The manufacturing process to create specific meat substitutes requires binding mycoprotein cells with other ingredients so that the muscle tissue structure of meat is simulated. Food technologists mix vegetable flavorings and egg white with mycoprotein to manufacture food products that appeal to consumers. They designed the mycoprotein-binder mix to resemble cuts of meat, patties, or nuggets. Those shapes are steamed to set the binder, then frozen and packaged for distribution. Marlow Foods Ltd produces Quorn wedges by fermenting wheat and corn sugars and mixing them with thickening agents and egg whites. The porous white substance can absorb spices, sauces, and flavorings for specific applications. Quorn Foods Inc. markets mycoprotein products as ingredients for cooking, including stir fry tenders, or complete entrées such as lasagna. These items can be grilled, cooked, or baked in regular kitchen appliances, including microwaves. They are intended to make food preparation convenient and quick.

Although protein shortages that had been projected in the 1960s did not develop, the early twenty-first century bovine spongiform encephalopathy (BSE) epidemic in Europe raised concerns about meat consumption and increased demand for alternative foods. For the most part, consumers interested in meat substitutes have accepted Quorn, and it became the leading synthetic food in Britain and international markets. Both health food stores and mainstream groceries in Europe and the Americas stock Quorn products. Approximately, one billion units of Quorn food were sold to an estimated 20 million people by the early twenty-first century. Besides its nutritional benefits, Quorn has been proven to reduce cholesterol levels and help dieters because people feel full after consuming Quorn products and are less inclined to indulge in excessive calories. Initially Quorn was expensive, but manufacturers have reduced prices to be similar to those of meat-substitute competitors.

When Quorn foods became available in the U.S. in January 2002, the Center for Science in the Public Interest (CSPI) demanded that the Food and Drug Administration (FDA) ban it. CSPI spokesman Michael F. Jacobson, PhD, blamed Quorn for U.S. consumers' digestive problems and attempted to brand it as a controversial new food. European consumers had not experienced similar Quorn-related health issues. Journalists reported that the CSPI had exaggerated and manipulated statistics, attributing symptoms from other disorders and causes to Quorn in an attempt to discredit the product. Resarchers initiated studies, including one at London's Royal Brompton Hospital, to evaluate Quorn and determined that it posed no significant allergen and health risks. Scientists consider mycoprotein among the safest proteins for dietary use. Because of CSPI pressure, the FDA did insist that Marlow Foods Ltd change labels inaccurately stating Quorn was derived from a mushroom family member because the company was afraid the term fungus would repulse consumers. Using chemostats, microbiologists continue to study how the mycoprotein fungus has evolved and mutated, and they have isolated variation strains to evaluate their impact on mycoproteins. Although Quorn had been economically profitable, AstraZeneca divested Marlow Foods Ltd to Montagu Private Equity in spring 2003.

Hydrogenated Fats

Food technologists create hydrogenated fats by processing vegetable oils, consisting of glycerides and fatty acids, with chemicals to achieve certain degrees of hardening. Partial hydrogenation stiffens oils, while full hydrogenation converts liquid oils into solid fat. The hydrogenation process involves moving hydrogen gas through heated oils in vats containing metals, usually copper, nickel, or zinc. When the metal reacts to the gas, it acts as a catalyst to relocate hydrogen molecules in the oil to create different, stiffer molecular shapes. This chemical reaction creates *trans* fats. Saturation of fats in these synthetic molecules increases according to the degree of hydrogenation achieved.

Hydrogenation makes oil-based food more manageable; for example, margarine can be made firm enough to form sticks, and it also enhances flavor and extends freshness and shelf life. Hydrogenated fats are frequently used as ingredients for doughnuts, chips, crackers, cookies, french fries, and candy bars, which are popularly known as junk food and have minimal nutritional value.

Nutritional investigations have deemed hydrogenated fats unhealthy if consumed excessively because of their saturated fat content. Medical studies have linked hydrogenated fats to interference with essential physiological chemical processes such as metabolism and lipoprotein receptor functioning. Hydrogenated fats increase risks of cancer, heart disease, and obesity. Most experts agree that eating unsaturated plant and fish fats in moderation is safer than consuming hydrogenated fats and

advise people to avoid products containing *trans* fats, which are unnecessary for bodily processes.

See also **Food, Processed and Fast; Food Preparation and Cooking**

ELIZABETH D. SCHAFER

Further Reading

Akoh, C.C. and Min, D.B., Eds. *Food Lipids: Chemistry, Nutrition, and Biotechnology*. Marcel Dekker, New York, 1998.

AstraZeneca shucks Quorn. *Chem. Eng. News*, 81, 14, 2 June 2003.

Avoid the fat trap. *Better Nutrition*, 64, 27 August 2002.

Fumento, M. Food fight. *Forbes*, 170, 56, 11 November 2002.

Grady, D. What's in those nuggets? Meat substitute stirs debate. New York Times, 14 May 2002, F1.

Griffen, A.M. and Wiebe, M.G. Extracellular protease produced by the Quorn myco-protein fungus *Fusarium graminearum*, in *Batch Chemostat Cult. Microbiol.*, 143, 3007, September 1997.

Hirsberg, C. Quorntroversy. *Vegetarian Times*, February 2003, 13.

Hunter, B.T. *The Great Nutrition Robbery*. Scribner, New York, 1978.

Hunter, B.T. Make way for mycoprotein in U.S. food supply. *Consum. Res. Mag.*, 84, 24, 2001.

Jacobson, M.F. New foods, new questions. *Nutrition Action Health Letter*, 29, 2, May 2002.

Knight, N. Roberts, G. and Shelton, D. The thermal stability of Quorn pieces. *Int. J. Food Sci. Technol.*, 36, 47, 2001.

Mann, G.V. Metabolic consequences of dietary trans fatty acids. *Lancet*, 343, 1268, 1994.

Mossoba, M.M., *et al. Official Methods for the Determination of Trans Fat*. American Oil Chemists' Society Press, Champaign, IL, 2003.

Pyke, M. *Synthetic Food*. J. Murray, London, 1970.

Rapoport, C. The Brits invent a new food. *Fortune*, 129, 16, 1994.

Sharp, T. Quorn mycoprotein: a new food for better nutrition, in *Implementing Dietary Guidelines for Healthy Eating*, Wheelock, V., Ed. Blackie Academic & Professional, London, 1997.

Spiller, G.A., Ed. *Handbook of Lipids in Human Nutrition*. CRC Press, Boca Raton, FL, 1996.

Underwood, A. Bad crop of Quorn? *Newsweek*, 140, 59, 2002.

Utley, A. "Queen for Quorn probes novel foods. *Times Higher Education Supplement*, 8, 21 November 1997.

Useful Websites

Marlow Foods Ltd: http:/www.quorn.com

Center for Science in the Public Interest (CSPI): http://www.quorncomplaints.com

Synthetic Resins

Chemistry became particularly conspicuous in the twentieth century through synthetic polymers. They include resinous products that are converted into plastics, laminates, surface coatings, and adhesives. Polymers exist because carbon has the property of forming single and multiple bonds with other carbons. In 1922 Hermann Staudinger suggested that polymers were macromolecules. Despite initial opposition, his ideas were accepted from around 1930 and had a considerable impact on industrial developments. The theory and mechanism of the processes whereby small molecules, the monomers, join in repeating units to create giant molecules, the polymers, was established around 1930, following the studies of Wallace Hume Carothers at the DuPont Company in the U.S. He identified two processes, condensation and addition, that distinguish between the main types of products. This provides a useful means for understanding historical developments.

From around 1900, chemists, electrical engineers, and inventors sought out novel products to replace or supplement natural rubber and gutta percha. Most promising was the chemical reaction between phenol and formaldehyde. Leo Hendrik Baekeland, a Belgian who had emigrated to the U.S., carefully controlled the conditions and recognized the catalytic action of acids and bases. He perfected the process in 1907. His main product was a resin readily converted into the first of the thermoplastics, those that set hard and rigid. Baekeland set up the General Bakelite Company in Perth Amboy, New Jersey, in 1910. Another early inventor was Sir James Swinburne, in England, but his process was covered by Baekeland's patent.

The availability of rapid hot molding from 1930 was a boon to growth. Bakelite was used extensively in the electrical and automobile industries, and also for cigarette holders, teapot and umbrella handles, telephones, and cabinets for radio sets. Bakelite had one disadvantage—its color, a dark brown, which at best appeared as black. It was followed with condensation products that could be pigmented, the amino resins. Almost transparent, they were made from urea and formaldehyde, and were discovered by Edmund Charles Rossiter in England at British Cyanides in Oldbury, Gloucestershire, during the early 1920s. They caused quite a stir when displayed during December 1926 in the form of molded colored tableware for the Christmas season at Harrods, in London. Melamine resins, made by a similar process, followed in 1939, and were mainly the work of the American Cyanamid Company in New Jersey. Melamine resins found considerable use in

wartime as laminates for maps, molding powders for trays and so on, and after 1945, as kitchen and bar tops, associated with Formica.

Alkyd resins are condensation products from polyols, such as glycerol, and dibasic acids, such as phthalic anhydride. First known as glyptals, they were introduced commercially in 1929 by General Electric in the U.S., mainly as a result of research by Roy H. Kienle. By 1938, production of alkyd resins exceeded the production of phenolic resins, since the greatest demand for resins was as coating materials.

A variety of resin formulations in which monomers containing the vinyl grouping, $CH_2{=}CH–$, became available from the late 1920s. These addition polymers include poly- vinyl acetate, -styrene, -vinyl chloride, -methyl acrylate, and -methyl methacrylate. In contrast to Bakelite and the amino plastics, their products are thermoplastic, which means that they can be softened and hardened by heating and cooling. Vinyl acetate resins were introduced in 1928 in the U.S. and Canada. Next was polystyrene, following the research of Hermann F. Mark and colleagues at IG Farben in Germany, and introduced commercially in 1932. Polyvinyl chloride (PVC), had been discovered at Griesheim-Elektron, in Germany, in 1912, but commercialization was delayed due to the problem of dealing with the hydrogen chloride gas that was evolved in the process. In the 1920s, B.F. Goodrich in the U.S. overcame this difficulty through the addition of plasticizers, though these were not required in the processes developed in the U.S. and Germany in the early 1930s. PVC was successfully introduced during 1932–1934, and was the most important vinyl product.

Acrylic is the generic name for polymers made from methyl acrylate, acrylonitrile, etc., though it is normally used with reference to methacrylates. The latter range from soft semifluids, used for adhesives and finishing of textiles, to hard, tough resins, for molded items and clear sheets. The possibility of an organic glass arose from studies at McGill University, Montreal, by William Chalmer. During 1928, the German company Röhm & Haas introduced a safety glass from methyl acrylate. In 1931, methyl methacrylate was found to afford a polymer that could be processed with greater ease. Polymethylmethacrylate (PMMA) became viable commercially in 1932, when John W. C. Crawford at ICI in England discovered an economical way of manufacturing the monomer. Production began in 1934, and the technology was licensed to Röhm & Haas, whose product was known as Plexiglas, in return for a license to manufacture cast sheet, ICI's Perspex, in Britain. The bulk of production until 1945 was employed in aircraft construction. Some was used to prepare false teeth, as a substitute for vulcanized rubber.

Carothers and colleagues at DuPont during 1928–1932 developed important routes to new polymers, including neoprene in 1929. Carothers discovered the condensation product nylon in 1934–1935, in part because he placed a strong emphasis on the structure of natural polymers. Nylon is a polyamide, as are proteins, made by condensation between 6–6 hexamethylene diamine and adipic acid. Introduced in 1938, nylon was the first commercial fully synthetic fiber. Famous for nylon stockings, its first use was in toothbrushes. Though Carothers' group also investigated the structurally similar polyester resins, prepared from an acid and an alkali, such as ethylene glycol and adipic acid, the first commercial product was discovered at ICI in 1942. This was the basis of the fiber Terylene, a linear polyester, polyethylene terephthalate. In 1942, glass-reinforced polyester was used in the manufacture of boats.

The discovery of polythene, however, took place in England. In December 1935 chemists at ICI investigated a white, waxy solid obtained when ethylene (ethene) was subjected to high pressure. It was found to be a polymer with excellent electrical insulating properties. The reaction took place because there was a leak in the apparatus, which allowed oxygen to enter, which then acted as a catalyst. By the end of 1938, one ton of polythene had been manufactured. A new plant was opened at the beginning of September 1939, just as World War II broke out. Polythene was used in high-frequency radio transmitters since, unlike other insulators, it absorbed little energy, due to its inert structure. From 1948 the flexible polythene became an important product used in consumer goods: bowls, buckets, wrapping film, carrier bags, soft drink, and milk cartons. In 1955 an underwater cable with polythene as the insulating material was laid down between Britain and the U.S. Polyethylene, normally familiar as a plastic, was later used in textile coating resins and emulsions.

The original process was worked at around 2000 atmospheres. In 1953, the German chemist Karl Ziegler at the Max Planck Institute discovered that polythene could also be synthesized at low pressures in the presence of somewhat expensive organo-metallic catalysts suspended in organic solvent. The process gave high-density polythene (HDPE), which was stiffer and well suited to the manufacture of crates. This process was subse-

quently adopted by Hoecht in Germany. Similar processes were investigated by Robert L. Banks and J. Paul Hogan at Phillips Petroleum in Bartlesville, Oklahama, who worked on polymerization of ethylene (1951), and propylene (1951–1953). There was considerable litigation over priority but the key patent was eventually awarded to the two Phillips' inventors. The Phillips product, called Marlex, was introduced in 1954 and later in the decade was used for hula hoop, a toy for children. Polymerization of propylene in the presence of a similar active catalyst, as discovered by Guilio Natta and co-workers at Milan Polytechnic 1954, led to the introduction of polypropylene in 1956. For their contributions, Zeigler and Natta jointly received the Nobel Prize in 1963.

Polyurethane contains the repeating urethane group, –NHCOO–. It was discovered by Otto Bayer, at Bayer in Germany in 1937, and developed commercially during the early 1950s in England and the U.S. It is a linear condensation polymer made from diisocyanate and a dihydric alcohol such as ethylene glycol, used in flexible and rigid foams, especially as insulation in fridges and freezers. Significantly, this is a condensation polymerization that, unlike most others, takes place without the loss of a small molecule.

Epoxy resins, such as the Swiss CIBA's Araldite, were discovered in the late 1930s, and introduced in 1946. They are made by condensation of epichlorohydrin and bisphenol A. The resins in the uncured state are thermoplastic, ranging from low-viscosity liquids to high melting-point solids. They are cured, or hardened, with polyamides, and used in surface coatings, adhesives, castings, and tooling applications.

See also **Adhesives; Composite Materials; Fibers, Synthetic and Semi-Synthetic; Plastics, Thermoplastics; Synthetic Rubbers**

ANTHONY S. TRAVIS

Further Reading

Bijker, W.E. *Of Bicycles, Bakelite, and Bulbs*. MIT Press, Cambridge, MA, 1995, pp. 101–197.
Blakey, W. The History Of Aminoplastics: The Sixth Chance Memorial Lecture of the Society of Chemical Industry. *Chem. Ind.*, July 1349–1357, 1964.
Fenichell, S. *Plastics: the Making of a Synthetic Century*. New York, Harper Collins, 1996.
Friedel, R. *Pioneer Plastics*. Wisconsin, Wisconsin University Press, 1983.
Furukawa, Y. Inventing Polymer Science: Staudinger, Carothers, and the Emergence of Macromolecular Chemistry. University of Pennsylvania Press, Philadelphia, 1998.
Hochheiser, S. *Rohm and Haas: A History of a Chemical Company*. University of Pennsylvania Press, Philadelphia, 1986. Methylmethacrylate resins.
Kaufman, M. *The History of Polyvinyl Chloride*. Maclaren, London, 1969.
Miekle, J.L. *American Plastics: A Cultural History*. Rutgers University Press, New Brunswick, 1996.
Morris, P.J.T. *Polymer Pioneers*. Center for the History of Chemistry, Philadelphia, 1986.
Mossman, S.T.I. and Morris, P.J.T., Eds. *The Development of Plastics*. Royal Society of Chemistry, Cambridge, Special Publication No. 141, 1994.
Seymour, R. B. *History of Polymer Science and Technology*. Marcel Dekker, New York, 1980.
Travis, A.S. Modernizing organic chemistry: Great Britain between two World Wars, in *Determinants in the Development of the European Chemical Industry, 1900–1939: New Technologies, Political Frameworks, Markets and Companies*, Travis, A.S., Homburg, E., Schröter, H. and Morris, P.J.T., Eds. Kluwer, Dordrecht, 1998, pp. 171–198.
Travis, A.S. Amino plastics and the melamine story. *Educ. Chem.*, 1, 16–19, 2000.
Vale, C.P. and Taylor, W.G.K. *Aminoplastics*. Iliffe Books, London, 1964. Includes a historical introduction, and a portrait of Hanns John who in 1918 patented the first industrial use of urea-formaldehyde resins.

Synthetic Rubber

Rubber is a ubiquitous material in modern society, enhancing the quality of life in a myriad of applications. It was unknown in the Western world until the Spanish began their explorations of America, where they found Indian tribes playing games with a ball made from the milky sap obtained by cutting the bark of local trees. In France this sap, or latex, was called caoutchouc after the native name for the "weeping tree" that produced it, while the English called it Indian rubber because it was useful for removing pencil marks from paper. A number of trees and shrubs produced such latex, including the orders Euphorbiaceae, Urticaceae, Apocynaceae, and Asclepiadaceae, but only two natural sources became commercially important (*Hevea brasiliensis* and *Parthenium argentatum*).

The use of rubber for practical purposes was slow to develop because the tree latex coagulated quickly and was difficult to process in the solid form. After solvents were discovered that would dissolve the solid rubber, products were made that took advantage of rubber's elasticity and waterproofing capability, but these crude materials suffered from an inherent stickiness and a form that changed depending on the temperature. With the discovery of the vulcanization process by Charles Goodyear in 1839, the rubber industry had a technique for eliminating these difficulties

and better consumer products soon appeared on the market.

Although plantations of rubber trees had been developed primarily in Southeast Asia, the demand for rubber became greater than the supply by the early twentieth century, and the rubber industry began a search for a viable synthetic rubber that would have the same desirable properties as the natural material. From the 1860s, chemists had been attacking the problem of the chemical composition of rubber, with the idea that once they could break natural rubber down into its chemical components, it would be possible to synthetically reconstruct rubber from those components and thus eliminate the need for the rubber tree.

Although this process has never been exactly achieved, it has been possible to prepare a large number of synthetic materials (mainly derived from crude oil) that have rubber-like or elastic properties. While they may be called "synthetic rubbers," they are not completely chemically equivalent to natural rubber, and are often called "elastomers" to reflect this difference. Thus, there is only one type of natural rubber, but there are many kinds of synthetic rubbers.

In 1860, Greville Williams in England isolated from natural rubber a relatively simple molecule named isoprene, a liquid hydrocarbon with the formula C_5H_8. On standing, pure liquid isoprene became viscous and formed a rubber-like material, leading to the exploration of catalysts that would enhance this process of polymerization, in which a simple monomer unit combined with itself many times to form a much larger molecule, or polymer. Chemists also examined the polymerization of other small monomer molecules with structures similar to isoprene, and were able to make a number of different elastomers, each with its own unique set of properties and its own advantages and disadvantages.

For example, one of the first commercial synthetic rubbers was methyl rubber, first produced by Bayer in Germany in 1910 from methyl isoprene and commercially produced during World War I. This was expensive and inferior to natural rubber, degrading when exposed to oxygen. Another synthetic rubber introduced in 1931 by Wallace Carothers at DuPont in the U.S. was neoprene, a polymer made from the monomer chloroprene, a derivative of isoprene that contained a chlorine atom. Neoprene was far superior to natural rubber in its resistance to gasoline and sunlight, and quickly found a market in the budding automobile industry. The question remained, however, if it would be possible to synthetically produce a general-purpose rubber with many uses, rather than specialty synthetics with limited applications.

The answer came in the development of copolymerization, a technique in which two different monomers combined in the polymerization process. Throughout the 1930s, several countries were seeking to reduce their dependence on natural rubber. In Germany copolymerization of styrene with butadiene (Buna S) or acrylonitrile (Buna N) had already been achieved by 1930 by IG Farbenindustrie in the laboratory. When World War II cut off the natural rubber supplies from the East Indies to the U.S., the U.S. embarked on a massive synthetic rubber program, deciding that the reaction of styrene with butadiene to produce a synthetic rubber similar to Buna S that the U.S. called GR-S would provide the best all-purpose rubber for the war effort. Since both of the monomers could be obtained from domestic sources, the U.S. began the construction of a number of plants for monomer and rubber production, and a new synthetic rubber industry was created where none had existed before. In 1939, U.S. synthetic rubber production was a meager 1700 tons. By 1943 production was 230,000 tons, and it reached 1,000,000 tons by the end of the war. In scope and complexity, this program rivaled construction of the atomic bomb in the Manhattan project.

After World War II, synthetic rubber plants were built worldwide, and by 1960 the use of synthetic rubber surpassed that of natural rubber for the first time. According to the International Institute of Synthetic Rubber Producers, by the end of 2001, "The yearly capacity of synthetic rubber manufacturing plants around the globe totals about 12 million metric tons and the capacity of tree-grown natural rubber produced on rubber plantations is approximately 8 million metric tons."

The Rubber Manufacturers Association indicates that about 70 percent of all rubber used today is synthetic in origin. There are about twenty different types of synthetic rubber currently in use, made from materials derived from petroleum, coal, oil, natural gas, and acetylene. However, there are a plethora of variations for each type (see Table 2). Where copolymerization is involved, both environmental conditions and amounts of the different monomers can be varied. It is also possible to blend different types of rubbers with each other and with other materials. Thus, the rubber industry can produce a vast number of

Table 2 Types of synthetic rubbers and principal uses.

Type	**Principal use(s)**
Styrene–butadiene (SBR)	Tires
Isoprene (IR)	Footwear, tires,
Polybutadiene (BR)	Blending with other rubbers in tires
Nitrile–butadiene (NBR)	Hoses, artificial leather
Acrylate (ACM)	Gaskets, textile coating, paper making
Chloroprene (CR)	Hoses, sealants
Chlorsulfonyl polyethylene (CSM)	Anti-corrosive coatings
Fluorocarbon (CFM)	High temperature seals and hoses
Isobutene–isoprene (IIR)	Tires
Ethylene–propylene (EPDM)	Automobile components
Ethylene-Vinyl Acetate (EVAC)	Cable coverings, textile proofing
Silicone (SI)	Aerospace, food processing, medical
Polyurethane (Ue)	Fabric coatings, insulation, foams
Thermoplastic Rubbers (TR)	Footwear and adhesives
Polysulfide (T)	Sealants for runways, bridges, structures

rubber products, each with its own set of unique properties that are required for a specific use. For example, the GR-S rubber of World War II, now called SBR rubber, is still the major synthetic being produced today, primarily because its properties are close to that of natural rubber; but there are over 500 different grades of SBR, which accounts for approximately 60 percent of total synthetic rubber production.

See also **Plastics, Thermoplastics; Plastics, Thermosetting; Synthetic Resins**

JAMES J. BOHNING

Further Reading

Bisio, A. *Synthetic Rubber: The Story of An Industry*. International Institute of Synthetic Rubber Producers, Houston, Texas, 1990.
Herbert, V. and Attilio B. *Synthetic Rubber: A Project That Had to Succeed*. Greenwood Press, Westport, CT, 1985.
Morris, P.J. T. *The American Synthetic Rubber Research Program*. University of Pennsylvania Press, Philadelphia, 1989.
Stern, H. J. *Rubber: Natural and Synthetic*. Maclaren, London, 1967.
Whitby, G. S., Ed. *Synthetic Rubber*. Wiley, New York, 1954.

Useful Websites

International Institute of Synthetic Rubber Producers: http://www.iisrp.com/synthetic-rubber.html
The Story of Rubber: A Self-Guided Polymer Expedition: http://www.psrc.usm.edu/macrog/exp/rubber/menu.htm

Systems Programs

The operating systems used in all computers today are a result of the development and organization of early systems programs designed to control and regulate the operations of computer hardware. The early computing machines such as the ENIAC of 1945 were "programmed" manually with connecting cables and setting switches for each new calculation. With the advent of the stored program computer of the late 1940s (the Manchester Mark I, EDVAC, EDSAC (electronic delay storage automatic calculator), the first system programs such as assemblers and compilers were developed

and installed. These programs performed oft repeated and basic operations for computer use including converting programs into machine code, storing and retrieving files, managing computer resources and peripherals, and aiding in the compilation of new programs. With the advent of programming languages, and the dissemination of more computers in research centers, universities, and businesses during the late 1950s and 1960s, a large group of users began developing programs, improving usability, and organizing system programs into operating systems.

Systems programs were developed for several early computers in the late 1940s and early 1950s. They developed along different lines but were responses to similar challenges that all early computer operators faced, namely, how to automate and make more efficient the programming of a computer. In 1948, David Wheeler, a doctoral student studying and working on EDSAC at Cambridge University, wrote one of the earliest but most vital programs called the "Initial Orders." This program allowed the computer, rather than a technician, to complete the tedious task of converting a symbolic sequence of instructions into the long strings of binary code that the machine could execute. Initial Orders was hardwired into EDSAC on rotary telephone switches before the main program was loaded. Later developments of the Initial Orders included a new innovation in system software called the Wheeler Jump. The Wheeler Jump facilitated the processing of short instruction sequences called subroutines during the operation of a main program. The EDSAC group at Cambridge continued this development of systems programs and in 1951 published the results of their research in their influential book, *The Preparation of Programs for an Electronic Digital Computer*.

In the early 1950s, Grace Hopper, one of the earliest American computer programmers, developed a program called the A–O compiler for the commercial computer manufacturer UNIVAC. The A–O compiler also attempted to automate the process of programming. Grace Hopper and her colleagues at UNIVAC, whose work paralleled that of the EDSAC programmers, also wrote, tested, and stored a library of frequently used sequences. The automatic programming concepts behind Hoppers' A–O compiler and the sequence library became integral components of systems software and allowed programmers to efficiently reuse code. Another notable early system program was developed in 1954. The program, a type of compiler that could convert user instructions into machine code, was installed on the Massachusetts Institute of Technology's Whirlwind computer. Although built specifically to handle algebraic problems, the system programs provided basic operating abilities such as converting and executing instructions and providing facilities for data storage and address location.

In the 1950s and 1960s, developments in systems programs also occurred among the users of computers. Assisted by the development of programs that aided programming, users had a greater range of tools to develop their own programs and modify their systems. One such user group, formed in 1955, was made up of users of the IBM 704 computer. The group, named SHARE, shared tips, developed programs, and had a significant influence on the development of several components for IBM operating systems. Also during the 1950s, the computer users at General Motors Research Laboratories devised a program to handle batch processing. This program, sometimes called a monitor, represented an early but successful system to manage and schedule a computer's resources.

As the 1960s progressed, user needs from the commercial, military, and academic sectors became more diverse and system software became increasingly complex, occupying large portions of the computer's memory. The large computer manufacturers such as IBM, Honeywell, and RAND focused their software development on systems programs. Separately sold application programs, such as word processing and spreadsheet software, did not enter the commercial market until 1965. The so-called software crisis of the 1960s was a recognition of the difficulty of organizing and coordinating system programs into a workable operating system. The development of IBM's OS/360 was an example of this crisis; released in 1966, OS/360 was 18 months behind schedule and $125 million over budget.

The 1970s and 1980s saw a turn away from some of the complications of system software, an interweaving of features from different operating systems, and the development of systems programs for the personal computer. In the early 1970s, two programmers from Bell Laboratories, Ken Thompson and Dennis Ritchie, developed a smaller, simpler operating system called UNIX. Unlike past system software, UNIX was portable and could be run on different computer systems. Due in part to low licensing fees and simplicity of design, UNIX increased in popularity throughout the 1970s. At the Xerox Palo Alto Research Center, research during the 1970s led to the development of system software for the Apple Macintosh computer

that included a GUI (graphical user interface). This type of system software filtered the user's interaction with the computer through the use of graphics or icons representing computer processes. In 1985, a year after the release of the Apple Macintosh computer, a GUI was overlaid on Microsoft's then dominant operating system, MS-DOS, to produce Microsoft Windows. The Microsoft Windows series of operating systems became and remains the dominant operating system on personal computers.

See also **Computers, Mainframe; Software Application Programs; Software Engineering**

JESSICA R. SCHAAP

Further Reading

Brinch Hansen, P., Ed. *Classic Operating Systems*. Springer-Verlag, New York, 2001.

Campbell-Kelly, M. and Aspray, W. *Computer A History of the Information Machine*. Harper Collins, New York, 1996.

Campbell-Kelly, M. Programming the EDSAC: early programming activity of the University of Cambridge. *Ann. Hist. Comp*., 2, 7–36, 1980.

Ceruzzi, P. *History of Modern Computing*. MIT Press, Cambridge, MA, 1998.

Rosen, S., Ed. *Programming Systems and Languages*. McGraw-Hill, New York, 1967.

Weizer, N. A history of operating systems, *Datamation*, 27, 118–134, 1981.

T

Tankers, *see* **Ships: Bulk Carriers and Tankers**

Tanks

Despite some curiosity on the part of a handful of other nations, the development of the tank in the twentieth century was largely a British affair. Yet even Britain did not intentionally set out to develop it. Like many things, the tank was the result of other technologies being developed as well as a response to the dangers some of those very technologies presented.

The three things needed in order for the tank to become a reality were power, protection, and vision. The first of these, power, evolved slowly. James Watt's improved steam engine in the latter half of the nineteenth century seemed to offer the as of yet unrealized promise of armored travel. In 1883, Germany's Gottlieb Daimler's four-stroke gas-powered engine, mounted on a bicycle, moved armored vehicles further along. In 1899 British engineer F.R. Simms created the "War Car," a Daimler-powered motorcycle, boasting a machine-gun mounted behind an armor shield. A year later Simms showed off a more deluxe version of the "War Car," this one encased in armor, bristling with a cannon and machine-guns and mounted on four steel-tired wheels. The age of armored, mechanized warfare had begun.

Once the need for power had been at least temporarily satisfied, protection became the next concern. Advancement in weapons systems such as high explosives and the machine gun made survival on the battlefield near impossible without armored protection. In England, Henry Bessemer's steel production process quickly replaced dependence on iron. Manganese steel proved to be better than ordinary steel, but it soon gave way to nickel steel. Both manganese and nickel steel seemed to offer an immediate, albeit expensive response to the development of metal-piercing munitions.

However, solutions to the problems of power and protection only proved that without vision, the tank would never become a staple of modern warfare. Despite the encouraging developments, warfare in the early twentieth century seemed to remain dependent on the type of horsepower that required oats. World War I began a shift in thinking. The first modern war, with long-range artillery, machine guns, air power, and trench fortifications proved deadly to advancing troops. The result was either a horrendous loss of life or a frustrating stalemate. In an attempt to break the logjam, Allied commanders who had once doubted the value of armored vehicles began to use them. Initially, armored cars were used, but the poor roads and muddy ground which became known as no-man's land proved to be the downfall of these wheeled vehicles. The British Navy then stepped in, and with the help of the Royal Engineers produced the next generation of armored vehicles. The result was at that point an unimaginable contraption. It was a machine of geometric proportion with steel tracks on each side, coated with armor and loaded with cannon and machine guns. At first it was called "Centipede," then "Big Willie," and finally "Mother." The evolution from F.R. Simms' "War Car" was complete—the tank was born. In September 1916, British M-1 tanks went into battle. Weighing in at 28 tons, the tanks carried a crew of eight, half of whom fired the two 6-pound (2.7-kilogram) cannon and four machine guns. They spanned the 12-foot (3.5-meter) enemy

trenches without mishap, negotiated the barbed wire and shell holes, and terrified the Germans, who quickly surrendered. Tracked vehicles would now be a new and necessary weapon. By war's end, France, Germany and the U.S., as well as Britain, were using tanks. The U.S. bought hundreds of French-built Renaults, and placed many of them under the command of Lt. Colonel George S. Patton, Jr. Patton, a former cavalryman, proved the tank's value not only as a supporter of advancing ground troops, but also as an independent assault weapon.

In the years following World War I, tank development varied. Nations, tired of war and the financial obligations it implied, either halted or slowed research and development. Germany, barred from any military development as a condition of the Versailles Treaty, could only experiment with tanks captured from the British. The U.S. Army did away with any plans to create an independent Tank Corps, relegating it to infantry support. The same lack of vision was evident in France and Britain, which also handed the tank over to the infantry. Nothing epitomized the lack of interest in the tank better than the experience of American engineer J. Walter Christie. Christie had studied World War I tanks and subsequently created a more modern fighting machine. His most important innovations were a suspension system, which made them more mobile, sloped armor to minimize damage from anti-tank weapons, a more powerful engine, and a larger gun mounted on a rotating turret. The U.S. was not interested, so in 1931 he sold his design to the Soviet Union, who modified it and created the "T" series of tanks, the first wave of armor to be designated as main battle tanks (MBT), which could combine mobility, lethality, and survivability. The Soviet's first effort, the T-34, equipped with a 76 millimeter gun proved to be up to the challenge of German tanks when Germany invaded the Soviet Union in 1941. Despite the restrictions imposed upon it after World War I, German military thinkers realized that the tank could make their armies more effective. They created entire divisions of tanks, with motorized infantry in supporting roles. It would take several years of fighting and dying before the Allied powers developed tank technology to challenge Germany's Tiger, Mark, and Panther series. Each sported heavier armor and firepower than anything America, France or Britain could put into action. The Panther was widely believed to be the best tank of World War II. The experiences of American and British tankers facing and surviving superior German technology contributed to the campaign in those two countries to pursue advanced tank technology in the postwar world.

That pursuit continued despite the belief that tanks had become outmoded in the nuclear age. The Soviet Union and France, as well as Britain and the U.S. took the lead in producing main battle tanks with greater speed, better protection and deadlier firepower. In particular the addition of infrared technologies on Russian and British tanks made them more effective in night engagements. Britain's Centurion series became an extremely effective armored vehicle when it was introduced after World War II, and as it was phased out in the 1960s it was snapped up in large numbers by other countries. The Centurion was replaced by the Chieftain, which in turn gave way to the Challenger, which featured a 120-millimeter gun, laser range-finding, thermal imaging, and night vision. The U.S. would ultimately become a leader in main battle tank innovation, but progress was slow. The wars America found itself enmeshed in, such as Korea and Vietnam, featured terrain inappropriate for extensive mechanized warfare. Additionally, air power had become the key piece in the U.S. defense puzzle. The result was a series of U.S. tanks, the "M" series, which were big, slow, underequipped in terms of firepower, and fueled with gasoline, making them instant fireballs if hit in the right spot. The M-60 was the last in the line, described by one American officer as "an inferior tank, part of a tired, old, second rate series." Meanwhile, the Soviet Union modified the T-34, and eventually spun it off into the T-72. Despite its cramped crew space, which occasionally exposed them to dangerous moving parts, Soviet tanks were formidable vehicles, with sloped armor, low silhouettes, and an accurate weapons system. In Britain, engineers found that angling the layers of armor on a tank could deflect anti-tank rounds. That development, along with breakthroughs in targeting and firepower produced new generations of tanks in Sweden, Germany and Israel. In the early 1980s the U.S. introduced the M1A1 Abrams, which received its first battlefield test during the 1991 Gulf War, and proved equal to the challenge. It cruised smoothly at 66 kilometers per hour, supported by torsion bar suspension and extra long tracks. The air-cooled gas turbine engine ran on a number of fuels, and produced less smoke and noise. It was also under 3 meterstall, which made it a hard target to hit. Inside, the four-person crew had a thermal imaging system to find and track targets, a 120 millimeter main gun with a range of 2500

meters, and a ventilation system to counteract nuclear, biological or chemical warfare.

Although tanks were plagued with problems of power, protection, and a lack of vision about their use at the beginning of the twentieth century, and despite the massive advances in weapons of all kinds during the century, at the end the tank remained a vital instrument of warfare.

See also **Warfare**

JOHN MORELLO

Further Reading

Fleming, T. Tanks. *Invention Technol.*, 10, 3, 54–63, 1995.

Macksey, K. *Tank Versus Tank: The Illustrated History of Armored Battlefield Conflict in The Twentieth Century*. Grub Street, London, 1999.

Technology and Ethics

Technology is manifest not simply in the mechanical, chemical, and electrical achievements from the first third of the twentieth century (e.g., automobiles, airplanes, synthetic materials, drugs, radios, motion pictures), the creations of physics that dominated in mid-century (e.g., nuclear weapons and space flight), and the electronic and biological inventions of the last third of the century (e.g., computers and genetic engineering) but also in the manifold social influences and impacts of these and related products, process, and systems. Such implications range from the economic and cultural to the legal and political—from new forms of production and consumption to the development of distinct governmental regulations and agencies. Yet underlying all such responses—and, indeed, originally calling forth the technologies themselves—are cultural and ethical commitments. As Aristotle argues in the opening pages of his *Ethics*, all human actions arise from some vision of the good; the critical examination of such visions as are present in the Greek "ethos," which we now call culture, is what constitutes "ethics."

The Ethos of Technology

From its sixteenth century beginnings the momentous cultural commitment to distinctly modern forms of making and using has rested on a vision of technology as a way to achieve the good if not a form of it. At the beginning of the twentieth century, Leo Baekeland, the inventor of Bakelite—the archetype for a whole family of plastics whose protean slickness contributed a new molecular level of artifice to such devices as the telephone and record player—summed up the prevailing moral assessment of technology as doing:

> "more for the betterment of the race than all the art, all the civilizing efforts, all the so-called classical literature" [*Baekeland 1910, p. 40*].

He continues:

> "The modern engineer, in partnership with the scientist, is asserting the possibilities of our race to a degree never dreamt of before: instead of cowering in wonder or fear like a savage before the forces of nature, instead of finding in them merely an inspiration for literary or artistic effort, . . . he fulfills the mission of the elect and sets himself to the task of applying his knowledge for the benefit of the whole race" [*Baekeland 1910, pp. 38–39*].

Taking hold of a baton passed from the sixteenth through the nineteenth centuries, the first half of the twentieth sought to extend what has been called "a second creation of the world," technologically transforming reality from something to which human beings adapted into that which they designed in their own image. This view of technology as justified by its pursuit and realization of material welfare and human freedom was reiterated fifty years later by C.P. Snow (1959) in his famous "two cultures" lecture, where he castigated literary intellectuals as "natural Luddites" for their worries about the materialistic values inherent in the great, ongoing achievements of the scientific and industrial revolutions: improved health care, increased food production, and the democratization of education. The primary ethical justification of modern technology across the century was the conquest of nature and the promotion of humanization as the pursuit of freedom.

During the second half of the century, however, the new technological world itself came to be recognized as requiring its own adaptations. Freedoms were not themselves always free. Thus, although the 1900s opened with a confidence in and remained throughout deeply sympathetic to an almost unqualified faith in the moral probity of technology, the later twentieth century witnessed the emergence, even within the techno-scientific community, of a series of critical ethical questions addressed to technological humanism.

The Question of Dehumanization

One powerful statement of this criticism argued that technological change fostered a form of dehumanization insofar as it separated human begins from nature and tradition, and subordinated the rich variety of human experience to judgments of instrumental rationalism. An espe-

cially influential articulation of technology as dehumanizing, inherited as well from the nineteenth century, focused on the issue of alienation in manufacture. For Karl Marx, alienation was narrowly defined in terms of the loss of control by workers over the processes and products of their labor, a loss sponsored by rationalist divisions of labor and large-scale industrialization. Division of labor is, however, but a special case of the more general phenomena of technology being disembedded or torn asunder from culture. Prior to the modern period, traditional technologies were characteristically embedded in a human life world; that is, hedged about with mores and counter-mores institutions. In agriculture, the major sector for employment before the twentieth century, the planting and harvesting of crops, the slaughtering of animals, and the consuming of foods were embedded in, that is, part and parcel of, age-old religious and cultural rituals and taboos. In many so-called pagan cultures the practice of agriculture required sacrifices to the gods; Hinduism, Buddhism, Judaism, and Islam all limit the partaking of certain foods, from animals to alcoholic beverages, and dignified seasons and celebrations with particular food associations.

It is not even correct to describe the traditional relationship as one of cultural ends guiding technical means, because our contemporary means–ends distinction was conspicuous by its absence in the premodern web of life. Each human activity was folded over or implicit in others. With industrial production, however, the web was unraveled so that means–ends distinctions came to the fore, and technology as means was cut loose from any particular ends in order to be pursued and developed on its own to an unprecedented degree. The result is what sociologist William Fielding Ogburn termed "culture lag," as life-ways ran to catch up with the explosion of new products and processes being introduced into human experience. The often-felt "loss of control" of our attempts to control nature are a further expression of this disembedding, and the root impetus calling for ethical reflection on the new means being placed at the disposal of a plethora of human desires and intentions cut free from traditional constraints.

Take some examples from the fields of transportation and communication. The automobile and the airplane turned the geographic niches of culture into porous boundaries, first nationally and then internationally, while introducing into ethos itself a manifold of unexpected consequences. The deployment of the assembly line by Henry Ford in 1914 transformed a toy of the wealthy into a consumer product for the many, facilitating the mid-century phenomenon of freeway suburbanization, while contributing to the sexual revolution in the form of a dating bedroom on wheels and to a geologically significant build up of atmospheric carbon. The invention of the airplane by the Wright Brothers in 1903 likewise transformed warfare, diplomacy, business, and tourism in diverse and sometimes conflicting ways that continue to play themselves out on the global stage. The communications technologies of telephone, radio, motion pictures, television, and computers further reduced the experiential importance of that place and history in which culture was once exclusively embodied. The displacement of ethos from localized initiation rituals of birth and death or forms of dress and speech into communications media styles and mixes supporting a cafeteria of entertainment programming is a uniquely twentieth century phenomenon that prolongs the disembedding trajectory. One correlate of such broadscale disembedding was what conservative culture critics of technology such as Lewis Mumford described as a lowering of the standards of taste in material culture and the replacement of spiritual discipline with the pleasure principle.

Questions of Ethics

Ethical criticism of such massive cultural dislocations remained more or less marginal until the invention of nuclear weapons took the vague unease characteristic of conservative intellectuals and placed the ethics of one specific technology in the forefront of public discourse. After Hiroshima and Nagasaki, many scientists and engineers found their gut intuitions given voice by J. Robert Oppenheimer, the engineer–manager of the atomic weapons program, when he said "In some sort of crude sense which no vulgarity, no humor, no overstatement can quite extinguish, the physicists have known sin." (Others violently complained that Oppenheimer had no right to beat his breast in public for them.) As Albert Einstein summarized the situation in less religious but nonetheless equally dramatic words, "The bomb ... and other discoveries present us with ... a problem not of physics but of ethics."

World War II likewise confronted the human community with instances in which even the most humane techno-sciences, those associated with medicine and its professional ethos of care, had been bent and corrupted through unthinking subordination to base political agendas. German

and Japanese physicians performed medical research on patients that amounted to forms of torture, while developing chemical and biological weapons for use on noncombatants and combatants alike. As a result, the Nuremberg War Crimes Tribunal sought to establish new and stricter guidelines for the conduct of medical experimentation on human subjects, making free and informed consent a paramount requirement and applying the principle of distributive justice to the apportioning of any benefits from such research. Subsequent investigations disclosed immoral medical experiments not only among the enemies of democracy but within democratic regimes themselves, with medical treatments being withheld from minorities, as in the Tuskegee experiments on African Americans suffering from syphilis, and with soldiers and citizens being exposed to harmful doses of radiation, as in the nuclear tests at Nevada and in the South Pacific, all in the name of techno-scientific knowledge production or national defense.

Indeed, in the five post-World War II decades one can plot a series of ethical discussions, often initiated by techno-scientists attempting to create appropriate cultures of containment for new technological powers:

1. In the 1950s it was the ethics of testing nuclear weapons, leading to the Limited Nuclear Testban Treaty (1963), as well as debates about the ethics of nuclear deterrence policies.
2. In the 1960s developments in artificial intelligence began to challenge traditional views about the uniqueness of human thinking, and biologist Rachel Carson's *Silent Spring* (1962) exposed the ecological destructiveness of excessive pesticide use—the latter of which lead to establishment in the U.S. of what became an internationally influential government institution, the Environmental Protection Agency (1970). Subsequent arguments for renewed appreciations of the natural world on both anthropocentric and non-anthropocentric grounds inspired a whole new disciplinary discourse of environmental ethics, and eventually the idea of a World Charter for Nature (1984).
3. During the 1970s issues of environmental health merged with questions about how to allocate equitably costly high-tech medical devices and treatments to engender, in a dialogue between biomedical practitioners and ethicists, the field of biomedical ethics or bioethics. Questions about the safety of the first genetically engineered organisms prompted genetic engineers in the early 1970s to adopt a voluntary, worldwide moratorium on this technology, in order to establish appropriate protocols for its safe pursuit.
5. In the 1990s it was topics such as the loss of biodiversity, global climate change, and reproductive cloning that became major foci for ethical discussion and debate.

Practical Responses

Practical responses to this spectrum of techno-ethical issues can be observed at the level of the professional scientific and technical community and at the level of public decision making. One of the most remarkable features of twentieth century technical professions is the effort to formulate codes of ethics to guide their members in dealing with a host of potential ethical dilemmas. Not only has the medical profession progressively refined its ethical guidelines for the treatment of patients, but engineers have formulated codes that go well beyond the promotion of corporate loyalty or professional interests. As the century opened, there were no explicit codes of engineering ethics. When such codes were initially formulated in the 1910s they emphasized responsibilities to employers and clients. By the end of the twentieth century, however, it was customary for engineering ethics codes to inspire their members to hold paramount "the safety, health, and welfare of the public" in the performance of their technical tasks and, indeed, to educate the public about the risks as well as the benefits of engineering projects.

Stimulated in part by a number of high-profile engineering disasters such as the nuclear meltdowns at Three Mile Island (1979) and Chernobyl (1986), the Union Carbide chemical plant explosion in Bhopal, India (1984), and the loss of the space shuttle Challenger (1986) engineering professionals also sought creative ways to educate and enforce their new codes, to support whistle blowers, and to engage the public in establishing appropriate institutions for the monitoring and regulation of technology. As a result, engineering curricula in many universities now teach engineering ethics, and the Institute for Electrical and Electronic Engineers (IEEE), the largest professional engineering society in the world, gives an occasional award for Outstanding Service in the Public Interest. The American Association for the Advancement of Science (AAAS), the largest

interdisciplinary scientific society in the world, likewise has a standing committee on "Scientific Freedom and Responsibility" which gives an annual award and works to engage science in the protection of human rights.

The 1980s also witnessed the exposure of a number of cases of misconduct, particularly in publicly funded biomedical research. This resulted in explicit legislative and institutional efforts to develop clearer guidelines for the responsible conduct of research in areas such as scientific record keeping, authorship and peer review, the treatment of laboratory animals, conflict of interest, and intellectual property rights. In the U.S. the National Institutes of Health, the largest funder of biomedical research, started to require ethics education for all graduate students and special boards of scientists and nonscientists to approve all research protocols involving human participants.

One convergence across these levels of professional and political response to the ethical challenges of technology was to abandon any strict laissez faire attitude toward whether or not, and how, the new powers of technological medicine and scientific technology might properly be deployed. Within the technical community a consensus emerged to try to avoid either a techno-scientific or market determinism in which things were made and used simply because they could be made and used. The basic belief was that enhanced and extended technological power called for an enhanced and extended practice of informed democratic consent in a world that was in effect becoming one gigantic socio-technical experiment. As Kristin Shrader-Frechette (1991) argued at length, insofar as the deployment of technologies constituted social experimentation, public participation is required. Yet public participation alone is not enough. Democratic intelligence depends on more than effective linkage of technological development to public desires or values; it also requires ethical insights to help it make informed and intelligent decisions about which technologies to accept, which to modify, and which to reject.

Theoretical Responses

The philosophical response that aimed to increase the kind of insight needed when faced with a thicket of ethical challenges centering around nuclear weapons, chemical engineering, high-tech medicine, computers, climate change, and biotechnology proceeded along two paths. One path has been to attempt a global or holistic assessment of modern technology as a transformation of the human condition. Here the work of such twentieth century thinkers as José Ortega y Gasset, Martin Heidegger, and Hans Jonas may serve as leading examples. Jonas, for instance, argues:

> "Modern technology has introduced actions of such novel scale, objects, and consequences that the framework of former ethics can no longer contain them... No previous ethics had to consider the global condition of human life and the far-off future, even existence of the race. These now being an issue demands... a new conception of duties and rights, for which previous ethics... provided, not even the principles"

In response, Jonas proposes a "heuristics of fear" to heighten the imagination of worst-case scenarios and thus introduce into the dynamism of modern technology a cautionary modesty.

One policy instantiation of such modesty is illustrated by the European Union adoption of the precautionary principle: essentially the view that a new technology should no longer be considered innocent until proven guilty (the classic modernist stance), but dangerous until proven safe. The weakness of such global stance, however, is that it remains at odds with the prevailing ethos of enthusiasm for technology still emergent elsewhere in the global marketplace, is difficult to implement in a pluralistic society, and abstract from any residues of those traditional forms of life that might actually support it. Besides, in concrete policy debates it is difficult to know how safe is safe enough.

Another path has been to take on problems one at a time, responding in a more piecemeal and pragmatic fashion, adapting received forms of ethical analysis and reflection. The two major modern ethical frameworks are known as consequentialism and deontologism. In consequentialist ethics, the rightness or wrongness of an action is dependent on the goodness or badness of its consequences or results (John Stuart Mills' utilitarianism is an example), whereas in deontological ethics rightness and wrongness are perceived as independent values of certain actions (Immanuel Kant's categorical imperative is the archetype). For instance, a deontologist might argue that respect for human autonomy and dignity demands without exception free and informed consent from all human research participants, even if this might compromise research that has a good chance of developing beneficial therapies. By contrast, a consequentialist may want to look at particular cases and insist that informed consent be justified on the basis of good results. In both frameworks, attempts have been made to understanding con-

sequences and articulate rights so as to better encompass the powers of advancing technology—but not always with success.

The most common late-twentieth century proposal to enhance consequentialism centered on risk cost-benefit analysis. The problem for consequentialism is the prevalence of unintended consequences and complex risks, especially those of low probability and high magnitude of harm (such as nuclear disasters) or epistemological uncertainties (such as the anthropogenic dimensions of global climate change). The existence of such cognitive debilities led David Collingridge (1980) to describe what he termed the paradox of the social control of technology: in the early stages of a technology, when it would be relatively easy to modify its development, we seldom possess the knowledge to make rational decisions; by the time we have more experience and a better understanding of its consequences and risks, technological momentum has made control difficult, if not impossible. This paradox suggests the need to develop social institutions dedicated to proactive technology assessment and, whenever possible, the choice of more over less flexible technologies.

With regard to deontologism, there has been a consistent effort to isolate a few firm boundary principles, as with the obligation to seek informed consent for any human subject experimentation. The problem is that in a techno-scientific, pluralist democracy, all fundamental principles tend to rest on a minimalist public consensus more than on reasoned insight—a fact that tends to promote advocacy lobbying. Under such circumstances deontological boundaries get restricted to procedural rather than substantive issues; and even there, subject to modification by strong public opinion. In the last quarter of the twentieth century, for instance, a deontological prohibition on the use of unproven drugs was undermined by AIDS activists demanding treatment.

Conclusion

By the end of the twentieth century the enthusiastic commitment to technology as being good under virtually all circumstances had been qualified by a more nuanced faith and by diverse efforts to bring critical ethical reflection to bear on the opportunities and threats associated with the most rapid and expansive period of technological changes in human history. Two other complementary developments were efforts to model complex phenomena and the rise of interdisciplinary and social networks, as manifest most conspicuously in both cases with computers and the Internet. To some extent, ethical reflection coupled with modeling and communications networks can be read as responses to the cultural disembedding that has been the hallmark of modern technology. But insofar as technology has been disembedded not simply from culture but from nature as well, the long-term viability of this solution is unclear. The rich diversity of the planet, the product of eons of evolutionary interaction across multiple scales may simply be impossible to model in its fullness. Models necessarily simplify and leave things out and are not terribly good persuaders of public action. Indeed, the century ended with the emergence of ethical questions about scientific and technological modeling.

See also **Technology, Society and the Environment**

CARL MITCHAM

Further Reading

Baekeland, L.H. Science and industry. *Trans. Am. Electrochem. Soc.*, 17, 37–53, 1910.

Harremoës, P. *et al.* Eds., *The Precautionary Principle in the 20th Century: Late Lessons from Early Warnings*. Earthscan, London, 2002. Fourteen case studies of such issues as radiation, asbestos, CFCs, and "mad cow disease."

Jonas, H. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Translated by Jonas, H. and Herr, D. University of Chicago Press, Chicago, 1984.

Katz, E., Light, A. and Thompson, W.B., Eds. *Controlling Technology: Contemporary Issues*, 2nd edn. Prometheus Books, Amherst, NY, 2003.

Mitcham, C. and Mackey, R., Eds. *Philosophy and Technology: Readings in the Philosophical Problems of Technology*. Free Press, New York, 1972; paperback 1983. Section two highlights "ethical and political critiques." Includes a select, annotated bibliography.

Shrader-Frechette, K.S. *Risk and Rationality: Philosophical Foundations for Populist Reforms*. University of California Press, Berkeley, 1991.

Technology and Leisure

Leisure, as defined by the *Encyclopedia of Social Sciences* (1933), is the "opportunity for disinterested activity," that is, activity that is not defined by its utility. Particularly in the industrialized world, leisure is usually contrasted to work; "free time" outside gainful employment is leisure time. The other important dimension of leisure activity is that it is voluntary, freed not only from the constraints of employment, but also open to the whims of the individual. In this sense, leisure is closely related to recreation and entertainment.

The contrast of leisure to work emerged around the beginning of the twentieth century. Mass industrialization and the factory system had forever altered the conditions of labor, dislodged the personal identification of many laborers with their work, and associated the workplace with drudgery. The Industrial Revolution also revised the pace and schedule of this work; by 1930, labor laws had significantly reduced the length of the typical working week in Europe and the U.S. Reactions to these changes were both optimistic and pessimistic. Some, such as the American writer Archibald MacLeish sang the praises of mechanization, believing it would lead to a "civilization in which all men would work less and enjoy more" (*The Nation*, 8 Feb. 1933). Others focused on the "leisure problem," the notion that idleness of the working classes—whether due to leisure time or mass unemployment—would lead to social unrest and immorality. International attention to the complex of issues raised by the transformation of labor came to a head during the 1920s and 1930s through organizations such as the International Labor Organization (founded in 1919) and international conferences that focused attention on workers' spare time.

The concept of leisure in contrast to work time is thus historically conditioned. It has only prevailed since the beginning of the twentieth century. Until then, leisure had been more typically depicted as a form of privilege, as a reflection of the status of free men; it conveyed the opportunity to devote oneself freely to pursuits such as the pursuit of knowledge. The tradition of considering leisure as the basis for contemplation and wisdom went back as far as Aristotle. In the late nineteenth century, writers such as Thorstein Veblen and Paul Lafargue insisted on the social acceptability and, in the case of Veblen, the benefits of idleness and leisure. Veblen's *Theory of the Leisure Class* (1899) depicted the "nonproductive consumption of time" in social terms, arguing that it was not mere idleness, but time spent privately in order to produce "immaterial" goods by those who could afford to do so. Examples of the results of leisure were "quasi-scholarly or quasi-artistic accomplishments," cultural observances, physical exploits, and the like (explained in the chapter entitled "Conspicuous Leisure,"). In short, leisure for Veblen was a "requisite of decency" in society.

The characterization of leisure as time away from work stimulated the creation of spaces for the "business of pleasure" (a term coined by Peter Bailey). These included sites that were specially constructed for the immediate delivery of entertainment, such as amusement parks, racetracks, and theaters. With the expansion of railroad networks, rise of the automobile, and introduction of airlines and cruise ships, travel to local, regional, and remote destinations for recreation and tourism also became feasible. In the last third of the twentieth century, the accelerating development of computer and network technology pushed the virtual worlds created through networked environments, video and computer games, and virtual reality, to the forefront of entertainment options, with a commercial impact rivaling that of older media such as motion pictures. In each phase of the expansion of available leisure spaces, technology played an important role both in expanding access to hitherto unavailable spaces and in creating compelling entertainment experiences within these spaces.

These notions of leisure and the activities associated with leisure time have thus been dramatically transformed over the course of the twentieth century. Generally speaking, the role of technology has been two-fold in this transformation. First, the deployment of technology has altered conditions of work and the efficiency of labor, and mechanization and industrialization have created a new context for leisure. Shortened working weeks, for example, stimulated discourse on the "problem" of free time. Religious, political, and social leaders considered the nature and importance of leisure activities, while industrial psychologists reconceived leisure as reducing the physical and psychological stresses suffered by the work force, particularly in the industrialized West. The notion of leisure was transformed, no longer a privilege of the "leisure class" but a space of possibilities for time away from work; something no longer to be repressed and feared, but cultivated. As Ida Craven asserted in an article on leisure for the *Encyclopedia of Social Sciences* in 1933, "the tone of any society is largely determined by the quality of its leisure." Technology has also played a second role in the shaping of modern leisure, less in the context than in the nature of leisure. Technological innovation and new systems of production have also transformed the practices associated with leisure time, creating for example a wealth of new recreational activities and entertainment media. At the end of the twentieth century, the appeal to consumers of free-time pursuits such as motion pictures, sports, computer games, and amusement parks depended more often than not upon the introduction of cutting-edge technologies into these realms. While a century earlier, technology, engineering, and factory production were

opposed to the contemplative pursuits of the leisure class, by the end of the twentieth century such diverse concepts as Walt Disney's "imagineering," the "studio system" in movie production, or videogame design suggested that entertainment and leisure had been fully integrated into post-industrial patterns of production and consumption.

See also **Changing Nature of Work; Computer and Video Games; Entertainment in the Home; Sports Science and Technology; Technology, Arts and Entertainment**

HENRY LOWOOD

Further Reading

Bailey, P., Ed. *Music Hall: The Business of Pleasure*. Open University Press, Milton Keynes, 1986.

Biggart, N. Labor and leisure, in *The Handbook of Economic Sociology*, Smelser, N.J. and Swedberg, R., Eds. Princeton University Press, Princeton, 1994.

Craven, I. Leisure, in *Encyclopedia of Social Sciences*, Seligman, E.R.A., Ed. Macmillan, New York, 1933.

Cunningham, H. *Leisure in the Industrial Revolution c.1780–1880*. Croom Helm, New York, 1980.

MacLeish, A. Machines and the future. *The Nation*, 8 February, 1933.

Veblen, T.B. *The Theory of the Leisure Class*. Macmillan, New York, 1899. Reprinted with an introduction by Lekachman, R. Penguin, London, 1994.

Technology, Arts and Entertainment

Introduction and Survey

Technology plays an important role in many cultural activities, especially in entertainment and in the fine and popular arts.

In cultural activities related to technical media several transformations—"mediamorphoses"—can be distinguished. In the graphic mediamorphosis of early modern times, messages and communications were transformed into print and music into notations; later, in the second half of the eighteenth century, chemical and mechanical mediamorphosis stored visual or aural realities. With photography, film, the wax cylinder, and audio disks, only recording and playback were required. It was no longer necessary to add a symbolic intermediate stage as in writing, printing, or musical notation. Laborious "true to nature" painting was largely replaced by the camera.

In the next stage, the electronic mediamorphosis of the first half of the twentieth century, different codes of communication were translated into electronic codes; the storage and playback of visual images and of sound played a dominant role with electronic recording replacing acoustical recording. With digitization in the 1980s, this development gained fresh momentum in a digital mediamorphosis. Electronic media for storing information were combined with digital computers, which provided increased storage and networking. With digital codification it became easier not only to reproduce "reality" but also to create "new realities"—computer-animated images for example. As a consequence, the difference between "pure" and "applied" art has become smaller and a new definition of art will soon be required. Whereas in the entertainment industry there has been a trend towards mergers and concentration, self-employment has likewise increased since the early 1980s. Together with large, expensive sound, film or TV studios, digitization has made the rise of small "bedroom productions" studios possible in which inexpensive audio or video productions can be made in a professional manner.

These activities described previously often interfere with copyright laws. Copyright has always been at war with technology because technical change often challenged intellectual property regulations and the accepted notions of authorship. New technologies of reproduction such as the photocopier or the video cassette have made it increasingly difficult for copyright holders to control the copying and distribution of their intellectual property; digital technology has added another dimension to this. Works in digital form are easy to copy and to distribute via the Internet; illegal use is difficult to detect and to prove. In line with this, the self-understanding of many modern artists equipped with scanners and samplers has changed significantly. They do not regard themselves any longer as "original geniuses" but as processors of information or manipulators of discovered material; a DJ's criterion for expertise is his skill in collating existing materials. There is a host of lawsuits pending on breaches of copyright, and in many cases pragmatic compromise arrangements have been made by giving licenses at the cost of a share of royalties.

Visual Arts and Technology

The history of art is, to a large extent, also a history of technology. Every art genre employs technology of some kind, the materials of visual art or the instruments of music making. There has been, and still is, continuous research on chemical processes for the development of paint varnishing that has affected painting in complex ways. After the introduction of monocular perspective painting

by the fifteenth century Italian painter Filippo Brunelleschi the invention of the camera obscura also played an important role. Similar to this significant change in the style of visual representation and the theory of visual perception, the invention of rapid serial photography by Etienne-Jules Marey and Edweard Muybridge in the 1870s and 1880s brought about a change in the way we perceive elements of visual motion and their representation.

In the early twentieth century the Italian futurists were fascinated by the machine and the motion of speed. Futurist esthetics valued the dynamic over the static and technology over nature. The cities with their noise and industrial rhythms were their preferred new material of art. Like the Italian futurists the Russian constructivists of the decade following the October Revolution of 1917 had an almost unlimited faith in the powers of industrial technology. Unlike the futurists, however, some of whom sympathized with Italian fascism, they dedicated their art to the service of the (Soviet) State. For the constructivists, the artist was an artist–engineer who made use of art in a radical reconstruction of society; but not all art movements of the early twentieth century were so fascinated by technology. Partly due to the terrors of World War I, which was also called the "war of the engineers," many artists realized technology's destructive power. The dadaists greeted technology with irony and the surrealists alluded to the disquieting and frightening aspects of the machine.

The introduction of electrical engineering and applications such as electronic means of controlling movement evoked the interest of artists. This gave rise to the "Kinetic Art" movement which investigated the dynamics of motion and the way, motion affects visual perception. Already in the late 1920s the theorist–artist Lászlo Moholy-Nagy, associated with the Bauhaus movement, which developed far-reaching insights into the relationship between technology and the arts, advocated a new perception of art in which dynamic values replaced static ones.

The Bauhaus movement in Germany in the 1920s proved enormously influential in all parts of the world. In a way it was a model for the art and technology movement in the United States in the 1960s. In the early 1960s, Billy Klüver, a research engineer and specialist in laser technology, worked together with artists like Merce Cunningham, Claes Oldenburg, Jasper Johns and Robert Rauschenberg, not only by providing technical assistance but as equal partner in the creative process. Collaborations with artists also led other scientists and engineers to make discoveries and achieve technical breakthroughs. However, the success of this cooperation was rather limited; many of these collaborations were marked by tensions and misunderstandings. For most of the artists involved, redefining the function of art was probably too radical a step. The same applies for the scientists and engineers, because members of those communities were rather wary of artists in general. Still, since the 1960s many industrial corporations with artists-in-residency schemes have provided artists with an opportunity to become acquainted with research and development activities in industrial companies, an arrangement that has proved beneficial to both parties.

Consumers and media saturation are themes highlighted by the pop art movement of the 1960s in which artists like Robert Rauschenberg (who also participated in the art and technology movement) and Andy Warhol played a large role. Pop art abandoned the artistic principles of modernism and related directly to the image-making technologies of mass culture. In Andy Warhols's machine esthetic there was a subversive effect. His technique featured the repetition, banality and boredom of automated production. This subversive complicity may also be observed in the works of many postmodern artists who make use of media technology.

In its fragmented, multiple narratives, multimedia performance exemplifies postmodern esthetics. It draws on a variety of different mixed media movements: the *avant garde* events of futurism and dadaism, the mixed media works of the 1960s fluxus movement; for example, Nam June Paik's sculptures made from television sets, as well the media-based interventions of the situationist movement. From the 1980s onwards, artists such as Laurie Anderson have incorporated digital technologies into their performances. Anderson's use of sampling and other sound-treating effects have to be seen in the context of her exploring technology's effect on gender and subjectivity. The postmodern performance work of Heiner Mueller deliberately dehumanized characters—technology itself becomes a character in his productions. In Peter Wilson's work, video and projected images are involved in a staged universe in which actors carry out repetitive tasks.

An art form that relies to a large extent on modern technological media is holography, a technique using light waves to record an image of a three-dimensional object on a two-dimensional photosensitive plate. In 1968, Stephen Benton of

the Polaroid Corporation introduced a new form of whole light transmission hologram; both the vivid coloring and the clarity of the holograms had a great impact on artists interested in the new medium. By the mid-1970s the gradual development of holography resulted in different holographic phenomena; for example, animated portraits using pulsed laser, motion pictures ("multiplex") holograms, acoustical holograms, and facilities to mass-produce embossed holograms. However, the tension between highly developed technical skills and the artistic content has always been a problem in holographic art. Recently though, artists have managed to successfully explore its potential for communicating human and esthetic values. By creating a new, esthetically meaningful paradigm for holography the new generation of holographic artists abolished old stereotypes.

Music and Technology

Technology has always been inseparable from music. The moment man ceased to make music solely with his voice, technology entered the scene. Some social philosophers argue that technology and musical techniques, content and meaning, develop together dialectically. Although technology has always played some role in music, an increasing technologization took place during the twentieth century, and some genres of contemporary music seem to be completely dominated by technology.

Already in the early twentieth century, an acceleration of the role of technology in music can be observed—a new "machine music" came into existence, electronic musical instruments were developed, and music composers seemed to turn into sound researchers. Recording engineers acquired increasing importance and the rise of studio esthetics had a significant effect on the listeners' expectations in concert halls. Shortly before World War I the Italians futurists demanded the rejection of traditional musical principles and their substitution by free tonal expression. This led to the design of noise instruments, the *intonarumori*. With the futurists and others, "electricity, the liberator" became the slogan of the day. Instruments like Thaddeus Cahill's "telharmonium" (1906) tried to imitate the sound of a symphony orchestra, insinuating that the orchestra players might soon be made redundant. The player piano lifted musical performance out of concert halls and transferred it to private homes. In the early 1920s Leon Termen built his "theremin," an electrical instrument helping which the human hand seemed to conjure sound from the air. This process was based on obtaining audible frequency beats formed by the interference of inaudible high-frequency oscillations. Many inventors of that time tried to design an electronic organ. In 1934 the American inventor Laurens Hammond developed the Hammond organ, an instrument with ninety-one small tone wheel generators with harmonic drawbars placed above the keyboard to permit the mixture of different tones. The instrument—easier to play than a conventional organ—proved immensely popular, as did the electric guitar, which in the 1950s and 1960s was to become the most important instrument of pop music. The guitar's amplification started in the 1930s in response to guitarist's demands for their solos to be heard over the sound of big bands. It facilitated an expansion of traditional guitar solo techniques and allowed the implementation of new techniques resulting in new effects like sustained tone.

Shortly after World War II sound recordings, disks and audio tapes were essential for the origins of "concrete music" composed from altered and rearranged sounds from the environment. Many composers in the (analog) electronic studios were not satisfied with this, however. They aimed at producing new sounds by applying simple oscillators to generate electromagnetic waves, which could then be translated into pure sound. Already in the mid-1930s the Russian physicist Evgenij Sholpo had applied the principle of artificially synthesizing an optical phonogram to his "variophone." In 1945 the U.S. inventor John Hanert with his "electrical orchestra" attempted to give the composer control over the complete fabric of musical composition. A tone was broken down into its characteristics—frequency, intensity, duration or timbre. Hanert thus reduced music to its constituent elements and reassembled it into coherent musical structures. In the 1950s, the RCA engineers Harry Olson and Herbert Belar made great efforts to synthesize sound. The apparatus they designed was, however, cumbersome and expensive. In 1966, Robert A. Moog started producing his Moog synthesizer using transistors and the technique of voltage control. He devised oscillators controlled by the amount of voltage that would alter the volume, pitch or overtones of the sound.

Although voltage-controlled synthesizers could produce a large variety of sounds and were immensely popular, their timbral capabilities remained limited. Computers and the "digital revolution" of the 1980s remedied this. Already

in 1956 Lejaren Hiller and Leonard Isaacson at the University of Illinois had experimented with computer music and used calculated procedures to generate musical scores. A year later, Max Mathews at the Bell Telephone Laboratories in Murray Hill, New Jersey, produced the first computer-generated sounds. The first experiments in digital synthesis were made in the mid-1970s; the "synclavier," invented in 1977, constructed every sound from scratch. The sampling techniques of the 1980s enabled musicians to treat all sound as data: once sampled, anything could be reproduced and reshaped. In 1983, the establishment of MIDI, or musical instrument digital interface, enabled musicians to easily transfer digital information between different electronic instruments as well as between instruments and computers.

Does all this mean that the introduction of electronics and computers into music making during the twentieth century has led to an ever-increasing musical perfection and to an opening up of new creative fields for professionals and amateurs? Some technological optimists are of that opinion and many "art music" composers, too, regard the computer as a useful tool for enhancing their creative abilities and relieving them from routine work. Others come to a negative conclusion: they argue that the computer stifles artistic creativity, produces a trend towards uniformity, devalues intellectual and artistic skills and has brought about technological dehumanization.

All these controversies aside, the revolution in music making as a consequence of electrification and electronics has taken place in only a few music genres. "Art music" making has proved remarkably resistant to electronics and computers; many composers of "minimal music" even feel disturbed by electronic sounds and prefer their music "unplugged."

Apart from music making, recording has been of great interest for the development of music in the twentieth century. Although many music enthusiasts were fascinated by it, conductors often declined to make recordings because they objected to their poor quality and resented the cold atmosphere in recording studios inimical to artistic inspirations. With improved recording facilities, however, the situation changed. In the 1960s Glenn Gould, the Canadian pianist, regarded the recording studio as the center of music making, relegating live performance to the fringe. Indeed, popular music from the 1960s onwards is unimaginable without the vast array of electronic studio equipment in existence; and other forms such as jazz would have developed differently without it. This is because recordings captured improvisations, which are extremely difficult to write down; but recording has also influenced music making. Before the rise of sound recording, for example, most violinists used vibrato sparingly. Once sound recording had been introduced, vibrato adopted a compensatory role. It made it possible for violinists to overcome the limitations of early recording equipment, served to mask imperfect intonation and also helped project a greater sense of the artist's presence.

Technology, particularly means of transportation like trains, cars and airplanes, have been an often recurring theme in both "art music" and "popular music." Many composers of the early twentieth century wrote music to reflect a changing world. In the 1920s railways and particularly railway engines aroused the interest of many composers, as is documented in Arthur Honeggers *Pacific 231*, named after one of the fastest American railway engines of its time. In *Pacific 231* the composer successfully transformed features like speed, dynamics, and energy into the language of music.

At the turn of the twentieth century artists also perceived the recently invented airplane as an esthetic event with wide-ranging implications for artistic and moral sensibility. Even more than in railways, artists and musicians transformed the airplane into a spiritual creation. The age of the airplanes was supposed to bring about unlimited individual mobility, peace, and harmony, but already before World War I it became clear that the airplane could also be used for destructive purposes. The utopia of peaceful internationalism gave way to aggressive nationalism and the two world wars bear witness to the misuses of flight. All these different feelings are reflected in paintings and literature, but also in twentieth century musical compositions. The airplane in the American composer George Antheil's Airplane Sonata (1921) manifests itself in machine-like driving rhythms and insistent ostinatos. The German composers Kurt Weill und Paul Hindemith transformed Charles Lindbergh's first transatlantic solo flight of 1927 into a dramatic tone poem (1929) and in 1945 the Czechoslovak composer Bohuslav Martini wrote his *Thunderbolt P-47*, a spectacular piece of program music, as an homage to the victorious U.S. Air Force. At the same time the American composer Marc Blitzstein wrote his symphony *The Airborne* about the glory but also about the terror inherent in aviation. In the second half of the twentieth century space flight also became a theme in music with generally positive connotations.

Performing Arts and Technology

Technology has always affected the theater. The Ancient Greeks introduced scenic painting but also the *deus ex machina*, a wooden elevator that brought a character onto the stage straight from heaven. In the eighteenth century, with the construction of the National Opera in Paris, there were wing-supporting chariots, overhead rigging sets, and an elaborate system of wing-operated floor traps set between the chariot slots. Those operations were mainly operated by "manpower"—a sizable number of stagehands—as well as by horses. In the nineteenth century, powered prime movers took over this task.

In the modern theater there are a variety of mechanisms used to erect, position, and manipulate scenery: hoists, lifts and horizontal drives. Hoists raise or lower scenery from the stage penthouse, lifts bring up scenic elements beneath the stage floor, and horizontal drives slide platforms, flats, and large properties from the wings onto the visible part of the stage or from one area of the stage to another.

Lighting and lighting effects have been of special interest in the performing arts for a long time. The modern era of stage lighting started with the invention of a practical electric lamp by Thomas A. Edison in 1880; gas lighting, which had been used before, was quickly discarded. At the turn of the twentieth century, incandescent lamps were in general use for stage lighting. Shortly before World War I concentrated oil filaments made incandescent spotlight possible, which in turn gave rise to the further development of stagecraft. Much lighting equipment was developed for special effects, moving clouds, for rain or snow or for fire effects. The oldest effects projector, the "Linnebach lantern," dates from the World War I era. In this "scene projector," as the Linnebach lantern was also called, a concentrated light source is placed in a deep black box; a painted slide is placed on the side of the box left open. The design painted on the glass is projected against a drop onstage, greatly enlarged, at a relatively short distance. As no lens is used in a Linnebach lantern, the light source must be powerful and concentrated.

Around the mid-twentieth century new interest arose in the use of projections. At a Wagner music festival at Bayreuth, Germany in the 1950s, Richard Wagner's grandson Wieland reduced the three-dimensional scenic elements to their essentials and then flooded the stage with multiple, overlapping projected patterns. Shortly afterwards the Czechoslovakian designer Josef Svobada developed his concept of "visions on space." He massed three-dimensional screens and, with slide and films, created a montage effect. In the early twentieth century motor-driven dimmers controlled auditorium lighting, but they were cumbersome and expensive for stage lighting. In 1890 the first remote-controlled dimmer was used, beset with problems. A good half-century later, remote control worked satisfactorily as the result of the work by George Izenour from Yale University, who in 1948 developed a dimmer with a thyratron, a type of electron tube. A few decades later, computerization increased the parameters of what can be done to enhance theatrical illusion and to come close the Wagnerian ideal of "total theater." With computerization hardly anything has to be left to chance.

Remote control is a key innovation in twentieth century theater technology. Computers are part of every operation. They seem to be best at repetitive or standardized operations and repetition, which is necessary for grand spectaculars requiring the complex coordination of various scenic elements. The director, as the main creative artist in the theater, is in charge of all this; interpreting the script, guiding the designers, and also manipulating the performers in order to realize his artistic objectives. Today many actors and critics argue that computer-controlled theater has established conditions in which the director, as puppet-master, is in total command of what is happening. He or she sits at the console and coordinates the sound, adjusts the lightening, and shifts the scenery. The director manipulates the performer, who is bereft of spontaneity and required to conform rather than to create. This pessimistic assessment plays down the important role that some directors already had in precomputer times and also the amount of artistic freedom famous actors still enjoy today, but the problem hinted at above certainly exists.

See also **Audio Recording; Film and Cinema**

HANS-JOACHIM BRAUN

Further Reading

Benjamin, W. *Illuminations.* Fontana, London, 1970.
Bettig, R.V. *Copyrighting Culture: The Political Economy of Intellectual Property*. Westview Press, Boulder, CO, 1996.
Bijvoet, M. How intimate can art and technology really be? A survey of the art and technology movement of the sixties, in *Culture, Technology & Creativity: The Late Twentieth Century*, Hayward, P., Ed. John Libbey, London, 1991, pp. 15–37.

Braun, H.-J. Introduction: technology and the production and reproduction of music in the twentieth century, in *Music and Technology in the Twentieth Century*, Braun, H.-J., Ed. Johns Hopkins University Press, Baltimore, 2002, pp. 9–32.

Braun, H.-J. Movin' on: trains and planes as a theme in music, in *Music and Technology in the Twentieth Century*, Braun, H.-J., Ed. Johns Hopkins University Press, Baltimore, 2002, pp. 9–32.

Carson, C. Theatre and technology: battling with the box. *Digital Creativity*, 10, 3, 129–134, 1999.

Chadabe, J. *Electronic Sound: The Past and Promise of Electronic Music*. Prentice Hall, Upper Saddle River, NJ, 1997.

Coyle, R. Holography—art in the space of technology: Margaret Benyon, Paula Dawson and the development of holographic art practice, in *Culture, Technology & Creativity: The Late Twentieth Century*, Hayward, P., Ed. John Libbey, London, 1991, pp. 65–88.

Douglas, D. *Art and the Future: A History-Prophecy of the Collaboration Between Art, Science, and Technology*. Thames & Hudson, London, 1973.

Duckrey, T., Ed. *Ars Electronica. Facing the Future: A Survey of Two Decades*. MIT Press, Cambridge, MA, 1999.

Elm, T. and Hiebel, H.H., Eds. *Medien und Maschinen. Literatur im technischen Zeitalter [Media and Machines. Literature in the Technological Age]*. Rombach, Freiburg, 1991.

Fidler, R. *Mediamorphosis: Understanding New Media*. Pine Forge Press, Thousand Oaks, 1997.

Flender, R. and Lampson, E., Eds. *Copyright. Musik im Internet*. Kadmos, Berlin, 2001.

Guderian, D., Ed. *Technik und Kunst [Technology and the Arts]*. VDI-Verlag, Düsseldorf, 1994.

Hayward, P. Introduction: technology and the (trans)formation of culture, in *Culture, Technology & Creativity: The Late Twentieth Century*, Hayward, P., Ed. John Libbey, London, 1991, pp. 1–12.

Hughes, R. *The Shock of the Modern*. Thames & Hudson, London, 1991.

Izenour, George C. *Theater Technology*, 2nd edn. Yale University Press, New Haven, 1988.

Katz, M. Aesthetics out of exigency: violin vibrato and the phonograph, in *Music and Technology in the Twentieth Century*, H.-J. Braun, Ed. Johns Hopkins University Press, Baltimore, 2002, pp. 174–185.

Kurzweil, R. *The Age of Spiritual Machines*. Viking Penguin, New York, 1999.

Lessig, L. *Code and Other Laws of Cyberspace*. Basic Books, New York, 1999.

Lunenfeld, P. *Snap to Grid: A User's Guide to Digital Arts, Media and Cultures*. MIT Press, Cambridge, MA, 2000.

McSwain, R. The social reconstruction of a reverse salient in electrical guitar technology. noise, the solid body, and Jimi Hendrix, in *Music and Technology in the Twentieth Century*, H.-J. Braun, Ed. Johns Hopkins University Press, Baltimore, 2002, pp. 186–198.

Manning, P. *Electronic and Computer Music*, 2nd edn. Oxford University Press, Oxford, 1993.

Millard, A. *America on Record: A History of Recorded Sound*. Cambridge University Press, Cambridge, 1995.

Murphie, A. negotiating presence—performance and new technologies, in *Culture, Technology & Creativity: The Late Twentieth Century*, Hayward, P., Ed. John Libbey, London, 1991, pp. 209–226.

Murphie, A. and Potts, J. *Culture and Technology*. Macmillan, Houndmills, 2003.

Pinch, T. and Frank, T. *Analog Days: The Invention and Impact of the Moog Synthesizer*. Harvard University Press, Cambridge, MA, 2002.

Rötzer, F., Ed. *Digitaler Schein: Ästhetik der elektronischen Medien [Digital Appearances: Aesthetics of Electronic Media]*. Suhrkamp, Frankfurt, 1991.

Schmidt Horning, S. From polka to punk: growth of an independent recording studio, 1934–1977, in *Music and Technology in the Twentieth Century*, H.-J. Braun, Ed. Johns Hopkins University Press, Baltimore, 2002, pp. 136–147.

Smudits, A. *Mediamorphosen des Kulturschaffens: Kunst und Kommunikationstechnologien im Wandel [Mediamorphoses of Cultural Creation:Change in Art and Communication Technologies]*. Wilhelm Braumüller, Vienna, 2002.

Theatrical Production, in *The New Encyclopaedia Britannica, Macropaedia*, 15th edition, vol. 28. Encyclopedia Britannica Inc., Chicago, 2003, pp. 561–615.

Théberge, P. *Any Sound You Can Imagine: Making Music/Consuming Technology*. Wesleyan University Press, Hanover, 1997.

Tisdall, C. and Bozzola, A. *Futurism*. Thames & Hudson, London, 1993.

Walker, J.A. *Art in the Age of Mass Media*. Pluto, London, 1983.

Technology, Society, and the Environment

The technology of the twentieth century significantly shaped not only society, but also environmental processes, in ways and on a scale unprecedented in human history. While the twentieth century "partnership" between technology and society in many respects continued nineteenth century trends, at least until the final decades of the twentieth century, the relationship with the environment changed fundamentally over the course of the century. Society for the first time significantly influenced, both advertently and inadvertently, not only the form but also the large-scale function of external nature. By the end of the century a thoroughly "technologized" society was uncertainly confronting the resultant challenges. Some of these, such as the risk of climate change and the resultant imperative to reduce greenhouse gas emissions, threatened to change the nature of industrial society itself (reasserting perhaps the more traditional tendency of the environment to shape society), while others, such as the challenge of regulating genetic technology, presented unparalleled ethical dilemmas. These developments are examined below through the lens of three overlapping clusters that exemplify the co-evolving relationship between technology, society, and

environment over the course of the twentieth century.

While many environmental developments, such as climate change, were global phenomena, the spread of twentieth century "technologized" society was not uniform. Centered in the leading industrial regions of the U.S., Europe, and Japan, twentieth century technology spread unevenly and at different rates across the rest of the globe. This account will concentrate on the leading industrial regions but within the context of this differential development. First a brief review of the emergence of significant interdependence between technology, society and environment during the twentieth century.

While nineteenth century railroads had significant social effects; for example, generating markets, helping impose rigorous conceptions of time and timekeeping, and facilitating the generation of national self-identity, many twentieth century technologies took such tendencies much further. Many technologies, including the automobile, electric light and telephone, deeply fashioned lifestyles and the twentieth century social order, influencing mental landscapes as much they did physical ones. By the end of the twentieth century it was common to think of technologies as "iconic," or emblematic of culture more generally. This increasing integration of society and technology mediated an unparalleled mutual dependence between society and the environment. By the end of the century social practices and behaviors, mediated by technology, were a significant factor in the world's climate and in the ozone concentration of the stratosphere, for example. Many related developments over the course of the century underline this interdependence. The role and scale of government grew with the necessity to support and regulate large-scale technological systems, such as those of transport and telecommunication, and also from the 1970s their environmental impacts. Trade and commerce, facilitated by advanced computing and communication technologies, took on a progressively more global character, reflecting the scale of environmental impacts; and dominant environmental attitudes evolved from regarding external nature as little more than a source of resources and sink for wastes, to a widespread concern with resource depletion and environmental pollution.

The first cluster to be considered, the "coketown cluster," emerged in the second half of the nineteenth century and centered on the coal, iron, steel, and railroad industries. These industries were concentrated in "smokestack" cities that bore the brunt of the urban air pollution that provided the environmental signature of this cluster. Originally centered in northwestern Europe and the U.S., it spread to Japan early in the twentieth century and the USSR in the 1930s, while later incarnations of it were still emerging in Soviet satellite countries and in China in the 1950s. This cluster was associated with a period of intense integration and consolidation in the world economy running from about 1870 until abruptly terminated by World War I. Facilitated and reinforced by emerging transport and communication technologies—the railroad, steamship and telegraph—and also by colonialism, it demonstrated many analogies with the globalization that marked the closing two decades of the twentieth century.

The dynamism associated with this burst in international trade and economic integration were enabled by contemporary technological developments such as refrigerated shipping, and bore witness to the voracity of the emerging urban markets of the smokestack cities. The environmental impacts of these developments started, for the first time, to become significantly disengaged in time and space from the activities driving them. For example the conversion of Argentinean pampas to beef production, New Zealand pasture to that of sheep, and Brazilian agricultural land to coffee production was driven by the pull of distant markets, whose environmental impacts became increasingly global. While consumer society, as it came to be known, was most intimately associated with the ensuing cluster, its essential ingredients were developing rapidly in the first decades of the twentieth century. By the beginning of the century even working class families throughout much of Europe and the U.S. could afford and had access to a growing range of consumer goods. Ready-to-wear clothing, bicycles, daily newspapers and comics, and standard forms of entertainment such as that of the earliest cinemas, signaled what was to become, in the following decades, a deluge (see below).

The growth of smokestack cities was facilitated, in part, by the extended reach of cities afforded by railroads and improved levels of urban sanitation resulting from the application of better wastewater and urban waste management technologies. The first decades of the century also witnessed the emergence of the modern city, driven in particular by innovations in electrical technology. One of these—the electric lightbulb—came to dominate street lighting throughout the industrialized world by the first decade of the century. The second was the advent of electrified tramways or trolleybuses.

Until these arrived (from 1890 onwards but increasingly from 1900) the spread of cities, facilitated by railroads, was constrained by transport from suburban stations being constricted to horse- and steam-powered vehicles. Electrified tramways or trolleybuses were faster, cheaper, cleaner, and more convenient and significantly encouraged the expansion of city boundaries. World War I, the first mechanized war in which tanks and aircraft played a significant part, signaled the end of the period of intense economic integration with which this cluster was initially associated. These various developments, however, heralded the emergence of the ensuing cluster that was to dominate the twentieth century, and one of whose signature technologies—the car—was to ensure that the age of electrified tramways and trolleybuses was short lived.

This next cluster, the "motown" cluster, was to dominate the twentieth century. At its core were the oil, electricity, automobile, aircraft, chemicals, plastic and fertilizer industries; and the powerful assembly lines whose facilitation of mass production paved the way for mass consumption, was one of its chief hallmarks. This cluster first took shape in the U.S. from about the 1920s, and quickly took hold in Canada, Western Europe, Japan, Australia and New Zealand and predominated from the 1940s. Its take-up was constrained by state ideology in the USSR and China, and by power and economic differentials in much of the rest of the globe, although by the end of the century this was changing with countries such as China rapidly superimposing motown characteristics onto still expanding coketown economies. While the motown cluster is complex, many of its essential characteristics are captured by the conjunction of Fordism, the name given to the principles of mass production, and the mass consumption that accompanied it. Henry Ford was pivotal to these developments. His first moving electrified assembly line (1912) and his seminal insight that wages should be set at a level that enabled workers to purchase the fruits of their labor; together with Taylorism (the idea that workers movements could be scientifically managed to optimize productivity), paved the way for consumer society. By 1923 Ford's workers could afford to buy a Model-T Ford with 58 days' wages. Fordism soon spread this principle to many consumer technologies, such as radios, washing machines and refrigerators, that became a hallmark of twentieth century industrial lifestyles.

One of the pivots upon which the growth of twentieth century consumer capitalism depended was marketing. As electricity grids extended rapidly in the early decades of the century it became clear that optimal use of the thermal generating plant that dominated supply provision required major loads additional to those of industry. Aggressive marketing, particularly of newly developed electric appliances, aided the rapid establishment of a significant domestic electricity market. Marketing was similarly a key feature of the rapid deployment of many seminal twentieth century technologies such as the phone and the car, upon which twentieth century lifestyles progressively depended.

The car exemplified these developments and was particularly influential in conditioning the physical form of cities, as well as lifestyles more generally. Most U.S. cities built after 1920 were shaped by the mass ownership of cars, a trend epitomized by Los Angeles where, similar to other U.S. cities, public trains were dismantled in the 1940s to make way for cars. It was at just this time that smog first became a political issue in Los Angeles. Photochemical smog, created by the action of sunlight upon the mixture of pollutants emanating mainly from cars, became a significant problem in many twentieth century cities, and is linked to a number of, principally respiratory, health problems. This, however, is only a fraction of the environmental impact of cars. Discussing these matters McNeill states, "In Germany in the 1990s [making a car] generated about 29 tons of waste for every ton of car. Making a car emitted as much air pollution as driving a car for 10 years. American motor vehicles (c.1990) required about 10 to 30 percent of the metals—mainly steel, iron and aluminum—used in the American economy. Half to two-thirds of the world's rubber went into autos." Cars also took up a lot of space and killed a lot of people (according to McNeill, from 25,000 to 50,000 per annum after 1925 in the U.S. alone) but remained perhaps the most coveted of twentieth century technologies.

Similar impacts were multiplied across the economy as a whole. The phenomenal scale-up in production and consumption facilitated by Fordism gave rise to immense increases in resource use and pollution (although improved technology made cleaner and less material and energy intensive production possible as the century progressed). Fordism's appetite for resources can be graphically illustrated with relevant statistics. Between 1900 and 1995, global coal output rose by a factor of over 6.5, while by about 1980, quarrying moved more earth than did natural erosion, for the first time. In simple physical terms this resulted in

major environmental impacts, many of the most significant and hazardous of which result from the wastes from these activities, in the form of stored wastes, slag, and tailings. Among other technological activities whose environmental impacts became particularly marked in this period was large-scale hydraulic engineering. Although dams had been used for electricity generation from the late nineteenth century the dynamics of the motown cluster inspired multipurpose (i.e., irrigation and power generation) hydro developments on a far greater scale. Pioneered by the Tennessee Valley Authority in the 1930s, and by the Boulder (later Hoover) dam on the Colorado river, the world's largest when built in 1935, large-scale dams came to appeal to twentieth century political leaders as emblematic of industrial progress. However, the enormous environmental and social dislocation brought about by large dams had made for a major rethink by the end of the century, except in China which was proceeding with the controversial Three Gorges Dam project, which if finished to plan will be the largest in history.

Many other technologies were significant in shaping technology, society, and environment relations throughout the twentieth century. Agriculture changed beyond recognition, not only via mechanization and motorization, but also crucially via the large-scale deployment of agrichemicals and of newly developed high-yielding crop strains. It was developments such as these that enabled previously marginal land, such as Australia's wheat belt, to be profitably cultivated so as to feed the vigorously expanding appetites of twentieth century cities. By the end of the century some of the environmental costs of such developments, such as the increasing salinization of agricultural land, which by the 1990s seriously affected about 10 percent of irrigated lands worldwide, were causing many to reflect on their wisdom. Also of particular significance was the mid-century development of nuclear technology that irrevocably changed the nature of warfare and whose civilian implications were still being calculated at the end of the century.

Other significant repercussions of the motown cluster included its effects on social arrangements, including those of a familial, gender, and economic nature. For example, while the traditional burden of housework was lightened considerably by the advent of cheap, readily accessable household appliances, women also became valued for many production line tasks in which precision and endurance counted for more than strength and skill. The renegotiation of the social contract implied by Fordism resulted in a significant reshaping of traditional class distinctions with a relatively leisured and affluent middle class emerging on a large scale in the U.S. from the 1940s, and a decade or so later in Europe. A paradoxical result of this emerging affluence was that it afforded the leisure to critically contemplate the status quo. One result of this was the rise of large-scale environmental concern in the 1960s and regulatory responses to it, most notably from the 1970s. Triggered by a concern over the impact of widespread pesticide use, environmentalism evolved, alongside fears of an energy crisis in the 1970s, into the emergence of ideas whose target became the form of industrial society itself. This emergence of environmental politics, a creature of the last three decades of the century, paved the way for and fed into the development of a new cluster.

The "postindustrial" cluster emerged in the 1980s and matured rapidly through the 1990s. It centered on the rise of sophisticated computing and telecommunications technologies (also called information and communication technologies, or ICT), biotechnologies, and, facilitated by ICT, a turn-away from primary manufacturing to a newly dynamic service sector. These both exemplified a technological trend to move away from the manipulation of material power to a focus on the manipulation of knowledge and information. The spread of this cluster was particularly uneven, both geographically and in terms of the adoption of the technologies it embodied. While centered in the traditional industrial heartlands of the U.S., Europe, and Japan, the advent of small, cheap, powerful distributed computing facilitated by the rapid extension of the Internet in the 1990s allowed many other countries to become involved, particularly in ICT. Malaysia, for example, while still industrializing (in both coketown and motown terms) had an ambitious national plan that placed it at the forefront of nations installing ICT infrastructure. Similarly in the 1990s, India became a center of software production, on which the rapidly expanding ICT depended. This was facilitated by the real-time access ICT granted overseas companies, such as those in the U.S., to the products of Indian software engineers, and by the relatively low level of Indian wages.

Contemporary commentators discussed how the scope and scale of the emerging ICT and the way they changed how time and space were experienced and conceived, heralded a new form of information or postindustrial society. The implications of these changes were unclear however by the end of the century. Many of the changes that were evident,

such as further deskilling of the workplace, simply intensified trends evident throughout the course of the twentieth century.

If anything, far greater uncertainties surrounded the biotechnologies emergent at this time. While promoted as having the potential to feed the worlds starving, the actual gains delivered by genetically engineered crops and foods by the end of the century, appeared to amount to little more than a continuation of those gains granted over the course of the twentieth century by more traditional selective-breeding techniques. The application of genetic engineering also presented both many ethical and moral dilemmas, and threatened a variety of environmental impacts that made many question its continuing expansion.

The tendency to critical evaluation marked by the emergence of environmental politics and ambivalence to emerging technologies was a hallmark of the final two decades of the century. From its beginnings in the 1960s, environmental politics was focused on the systemic environmental effects of industrial society, which both marked it out from an earlier conservation movement and made for a departure from the technological optimism that had dominated the century. Events only reinforced and confirmed these trends. Notable among these were the emergence of the problems of acid rain (see Electricity Generation and the Environment), ozone depletion and climate change. While the Montreal Protocol (1987), and subsequent amendments to it, most notably London (1990) and Copenhagen (1992), are widely regarded as having brought ozone depletion under control, climate change was another matter. The push to decarbonize the energy economy its control necessitated was a challenge many governments baulked at and by the end of the century the implementation of the Kyoto Protocol (1997) was uncertain and disputed (see Electricity Generation and the Environment).

Decarbonization was part of a broader reaction to the polluting and resource-intensive technologies that had dominated the century. A push to dematerialize industrial economies by reinforcing trends to greater efficiencies in production and further reduce the intensity of energy and resource use marked this reaction. Operating under the rubric of sustainable development, first popularized by the 1987 report of the UN Commission on Environment and Development entitled "Our Common Future" and legitimated by the 1992 Earth Summit in Rio de Janeiro, the 1990s were marked by a variety of approaches to these matters. While many of these, such as notions of clean (or sometimes cleaner) technology and that of industrial metabolism (which aimed to mimic nature by reusing and recycling waste streams) held great promise, there was some incoherence about their development.

While much of the reduction in the energy and resource intensity of leading industrial economies resulted from improved industrial efficiencies, a great deal of it, from about the 1970s, also resulted from the rapid emergence of a service sector. This occurred as traditional smokestack industries moved to less developed nations. The service sector, which is far less energy- and resource-intensive than traditional smokestack industries, centered on activities such as finance, tourism and the retail sector, and was given a boost by the popularity of economic deregulation during the 1980s and 1990s. In the mid-1990s the World Trade Organization was formed to extend economic deregulation to the sphere of global trade. This period in many ways picked up the drive to integrate the world economy with which the century started while also reflecting its laissez-faire economic outlook, but this time, aided by ICT, it was far more successful. Unfortunately, however, this expansionary success ensured that the push to dematerialize industrial economies was marginal to continuing global motown trends in resource and energy use.

It was with such tensions that the century closed. Underpinning these tensions were emergent signs, signaled for example by the Kyoto Protocol, of recognition of the need for a new contract between technology, society and environment; one in which technological development, informed by an increasing awareness of the intimate interconnections between technology, society, and environment, would be driven by debate and consent rather than simply by profit or serendipity. These, however, were matters for a new century.

See also **Electricity Generation and the Environment; Environmental Monitoring**

STEPHEN HEALY

Further Reading

Blunden, J. and Reddish, A., Eds *Energy, Resources and Environment*. Hodder & Stoughton, London, 1991.

Elliot, D. *Energy, Society and Environment*. Routledge, London, 1997.

Fichman, M. *Science, Technology and Society: A Historical Perspective*. Kendall/Hunt Publishing, Dubuque, IA, 1993.

McNeill, J. *Something New Under the Sun: An Environmental History of the Twentieth Century World*. Penguin, London, 2000.

Melosi, M.V., Ed. *Pollution and Reform in American Cities, 1870–1930*. University of Texas Press, Austin, 1980.
Roberts, G., Ed. *The American Cities and Technology Reader: Wilderness to Wired City*. Routledge, London, 1999.
Silverton, J. and Sarre, P., Ed. *Environment and Society*. Hodder & Stoughton, London, 1990.
Smith, D.A. M*ining America: The Industry and the Environment, 1800–1980*. University Press of Kansas, Lawrence, 1987.
Stradling, D. *Smokestacks and Progressives: Environmentalists, Engineers, and Air Quality in America 1881–1951*. John Hopkins University Press, Baltimore, 1999.
Tarr, Joel A. *The Search for the Ultimate Sink: Urban Pollution in Historical Perspective*. University of Akron Press, Akron, OH, 1996.

Telecommunications

The history of two-way communication is embedded in a series of much deeper histories: technology, imperialism, and the rise of the nation state, to name but a few. These, in turn, are embedded in a discourse that returns theory to a central place in non-Marxist histories. Several theoretical enterprises emerging from the academic disciplines of economics and political science look to technology as one of the major engines that drive history and to telecommunications technology as the most significant technology in the current world economy.

Tilly notes that there are essentially two types of states: trading states and territorial states.

Trading states are maritime in nature, exert weak geopolitical control over long distances, and excel at the long-distance communications needed to manage their economies. To exchange goods profitably, trading states have needed a great deal of mutual trust and sophisticated commercial arrangements among a community of spatially distant merchants. Updating Tilly's argument, one would add managers and consumers to merchants. Trading states have fairly fluid social structures and a mutually supportive relationship between commerce and government. In the political realm they tend to relative democracy. The archetypes have been Holland and Britain.

Territorial states are continental in nature, exert strong geopolitical control over short distances, but have very little reach beyond that. The economies of territorial states tend to be command economies, with the government in charge, and with a one-way information flow down a strict social hierarchy. In the political realm they tend to relative autocracy. The archetypes have been China and Russia. Some states such as the U.S., France, Germany, Japan, and Spain have caused untold chaos at different times in world history shifting between different aspects of the two types or, worse, attempting to be both at the same time.

From the1400s onward, wealth generation has been driven by the remarkable success of trading states and their ever-increasing domination of the world economy. During the 1900s, this proved to be even more the case, although it did not seem that way in 1904 when British geopolitician Mackinder suggested that nineteenth century improvements in communications, especially terrestrial telegraphy and railroads, meant the commercial domination of the world economy by trading states was ending. In the early 1900s, however, wireless telecommunications began to shift the balance back toward trading states. The three main wars of the twentieth century (World Wars I and II, and the Cold War) were about the containment of the two main territorial states of the period, Germany and Russia, by the two main trading states, Britain and the U.S. Much of this containment was through superior telecommunications.

Although the global telecommunications system was shaped by the spectacular growth of the British submarine cable system after the first successful transatlantic cable opened for business in 1866, the 1900s saw considerable improvement of a system that had reached much of its modern shape by the late 1800s. Seven innovations stand out:

1. Low-frequency wireless telegraphy, starting around 1900
2. High-frequency wireless, which allowed telephony as well as telegraphy, in the 1920s
3. Microwave wireless, which led to radar, in the 1930s
4. Submarine telephone cables in the 1950s
5. Satellites in the 1960s
6. Fiber-optic submarine cables in the 1980s
7. Cellular wireless telephones in the 1990s.

These seven innovations have increased the capacity or convenience of the global communications system and have trended towards radically lowered costs. Developments in two-way communications often encouraged one-way communications, such as the development of broadcast radio entertainment out of wireless telecommunications. Occasionally the flow of innovation has reversed, as in the importance of electronic television technology to the development of radar.

Two areas of interaction between commercial and state power have been driven by telecommunications. First, states have sought to control com-

munications systems, to prevent other states controlling them, or both. Most states, America being a rare exception, developed post, telegraph, and telephone systems (PTTs) as government monopolies in the 1800s, and frequently added such new technologies as broadcasting as they arose in the 1900s. Second, wireless telecommunication allows real-time connections without people or vehicles being tied to a spatially fixed infrastructure. The advantages of the cellular telephone to individuals are obvious. Commercially, the global positioning system (GPS) allows us to operate the world economy in an efficient, "just-in-time" way. Militarily, GPS allows considerable geopolitical control at very great distance, with the extreme accuracy and concentration of force afforded by "smart" bombs and cruise missiles, and without committing large numbers of ground troops.

Telecommunications are inextricably bound up with geopolitical power. In 1866 British capital began the successful installation of a global network of submarine cables that was never surpassed, investing as freely in Buenos Aires as in Birmingham. The geography of the network was such that most of the world's telegraphic messages passed through Britain. The cable companies involved cooperated closely with the British government. In World War I, Britain's control over global telecommunications and the skill of British code breakers was so complete that Room 40 in the British Admiralty was able to intercept and decode the notorious Zimmerman telegram. In 1917 German foreign minister Zimmerman offered Mexico the return of historic Spanish possessions in the American Southwest for attacking America, helping bring America into World War I.

The technological history of telegraphy is in three parts: terrestrial, submarine, and wireless, with only the last being a technology of the 1900s. The first great terrestrial companies were established in America after 1844. The first global telecommunications system was the British submarine cable system that grew after 1866, but required "repeater" stations on islands under British control dotted around the world's oceans, hence some of the odder territorial acquisitions of the Empire. Of the seven innovations in telecommunications of the 1900s, the first—low-frequency wireless—built on this base of global control.

Low-Frequency, Long-Wave Wireless

The possibility of wireless telegraphy was demonstrated by Hertz's empirical verification of Maxwell's equations in 1888. Eight years later, Marconi patented the first wireless telecommunications system. In 1901 Lloyd's gave Marconi's a monopoly on ship-to-shore wireless and required it for a ship to achieve their highest insurance rating. Marconi's commercial success was assured and the *Titanic* disaster of 1912 made Marconi a household name.

Wireless quickly became one of the most contested technological arenas of the early twentieth century. Nonmilitary researchers and entrepreneurs of this pioneer period dreamed of something like cellular telephones, well beyond the available technology, but it was in the geopolitical and military arenas that wireless took deepest hold. Several countries, the U.S. and Germany in particular, saw it as a way of challenging Britain's telecommunications dominance with all that that implied for British global hegemony. At the military level, navies quickly realized that wireless offered centralized command and control of ships at sea and began to force the ship-to-shore technology Marconi pioneered. Afraid of depending on a British supplier, the U.S. and German navies made major investments in wireless before 1910.

World War I caused even more rapid evolution of wireless than the naval armaments race before the war. By 1915, the British Army was using low-frequency wireless telegraphy to allow airplanes to spot for guns. By late 1918, Allied airplanes were being fitted with higher frequency, wireless telephone systems to allow centralized command and control of the air war. "Plan 1919" called for Allied wireless telephony to coordinate air and ground advance using attack airplanes and tanks. In World War II, Germany, laggard in 1918, demonstrated such *blitzkrieg* to perfection. The huge numbers of vacuum tubes produced for "Plan 1919" allowed commercial broadcasting to develop in the early 1920s, operating at the same frequencies as the radios built for mobile war, today's medium waves or AM band.

High-Frequency, Short-Wave Wireless

Vacuum tubes also made possible much higher frequencies for telecommunication. Marconi experimented successfully in the 1920s with high-frequency short waves, and "beam" antennae for global wireless telegraphy and telephony. The resulting imperial network of short-wave beam wireless stations was extremely cheap to build and operate and effectively renewed Britain's control of global communications by the mid-1920s. The submarine cable companies, seeing their profits

evaporate, persuaded the British government to save them in 1929 by forcing Marconi to merge with them. The resulting company became Cable and Wireless in 1934.

Microwave Wireless

Although microwave wireless had its origins in the commercial world when Standard Telephones and Cables, based on work in their Paris laboratories, installed a "micro-ray" telephone relay across the English Channel in 1931 operating in the 17.6-centimeter band, microwave technology matured in war, not peace, and with radar, not telephony. In 1908 Wells' novel, *The War in the Air*, suggested an irresistible fleet of German zeppelins could cross the Channel, destroy the British fleet, bomb Britain into submission, then continue across the Atlantic to destroy the American fleet and bomb New York. Zeppelins fared poorly in World War I, but German airplane raids scared Britain into developing the world's first independent air force to both attempt defense against bombers and to bomb back. Three strands of reasoning developed: in Italy, Douhet argued that bombing civilians into submission would be the way to win future wars; in America, Mitchell argued for precise bombing of strategic targets; and in Britain, Trenchard argued that air forces could control fractious provinces of the Empire more cheaply and with less loss of (British) lives than could occupying armies, a policy known as "control without occupation."

Despite arguments that "the bomber will always get through," defense began to seem possible when Watson-Watt suggested to Britain's Committee for Imperial Defence that radio might locate incoming bombers. Watson-Watt routinely transmitted bursts of high-frequency radio to examine the ionosphere and inform the imperial short-wave beam wireless stations what the best frequencies were that day. He noted early returns when aircraft flew overhead during transmissions. Radar matured very rapidly under renewed German aerial threat after 1934. Radar was not really new, being patented in Germany in 1904 to help ships enter harbors in fog. STC's "micro-ray" telephone link across the Channel led the French to equip Atlantic liners with experimental (and unsuccessful) collision-avoidance radar in the late 1930s. At the military level, after World War I all the major navies experimented secretly with radar to solve the problems of gunnery in poor visibility; by World War II all had implemented radar systems.

It was air war did the most to improve radar. In ships or onshore radar could be heavy, bulky, and use lots of energy. Britain's Committee for Imperial Defence decided by 1936 that a day battle for Britain could be won using ground radar-directed fighters, forcing Germany into night bombing and requiring lighter, smaller, more energy-efficient microwave systems in fighters themselves. Massive technology forcing ensued and the British deployed Mark IV airborne intercept radars operating at 1.5 meters in early 1941, during the nighttime blitz on London. In part, this technology succeeded by drawing on Britain's pioneering electronic television, commercially deployed in 1936. Television engineers maintained the crucial radar systems and, at war's end, radar engineers found employment in television. Blumlein, Britain's greatest electronics engineer, helped develop electronic television, Mark IV airborne intercept radar, and H2S ground-imaging, bomb-aiming radar before his untimely death in June 1942. H2S radar was crucial for Britain's commitment to strategic bombing, allowing bombing by night or through thick cloud or smoke. It and its counterpart, Airborne Intercept Mark IX radar, operated in the 10-centimeter microwave band, later moving to 3 centimeters.

After World War II was over, microwave technology returned to its peacetime origins in microray telephony. Because little real estate had to be acquired, just hilltop sites for line-of-sight relay towers, microwave telephony was cheap, offered high-capacity communications, and was rapidly installed in countries such as the U.S. and Canada where sheer distance made co-axial cable expensive. By the late 1940s such microwave relays were also planned to carry television signals coast-to-coast in North America.

Submarine Telephone Cables

In 1956 Bell Telephone laid the first transatlantic telephone cable, TAT-1, using analog transmission technology. Initial capacity was only 36 simultaneous calls but it also carried more telegraphic traffic than the then total capacity of all existing submarine telegraph cables, rendering them obsolete. Unlike the submarine telegraph cables, telephone lines need a great deal of power to drive speech long distances. Bell devised vacuum tube amplifiers compact enough to be installed within the cable and reliable enough for 25 years service on the sea floor. TAT-1 needed 51 amplifiers in each direction to push signals across the Atlantic.

Although such telephone service was well beyond the pockets of the average person, TAT-1, its successors through TAT-7, and its competi-

tors, the Anglo-Canadian CANTATs-1 and -2, ushered in a revolution in business communications across the Atlantic. Capacity mushroomed, from 36 simultaneous calls in 1956 to 11,173 when TAT-7 opened in 1983. Together with jets, which entered transatlantic service in 1959, the TATs made possible effective multinational corporations. The large American automobile manufacturers, Ford and General Motors, became proto-multinationals before World War II, but operated their American, European, and Japanese factories as independent fiefdoms with independent management structures. The communications and transport revolutions of the 1950s made centralized management possible from America through instant voice communication and rapid site visits to remote factories. As the European economy recovered following World War II, European companies such as Nestlé followed suit.

Pacific telephone cables and the development of jets able to reach Tokyo from London or New York nonstop have ushered in the era of global companies, especially once the Cold War ended and Russian airspace was opened to Western jets. Telecommunications capacity grew more slowly across the Pacific in the analog era. The first Pacific telephone cable was Transpac-1, installed in 1964 with 142 lines. By 1977 there were only 987 lines from America to Asia with a further 1380 from Canada to Australia and New Zealand.

Submarine telephone cables had severe limitations. TAT-1, at approximately \$1 million per voice circuit, made calls expensive. Even TAT-6, at \$179 million for 4,000 circuits, was stunningly expensive, as was TAT-7. AT&T proposed a 16,000-circuit analog TAT-8 that would have cost close to \$1 billion and required a thousand built-in amplifiers. A cheaper solution was needed.

Satellites

Satellites were seen as the first alternative to the expensive, low-capacity submarine telephone cables of the 1950s. The first communications satellite, Telstar, was designed by Bell Telephone and launched into low-earth orbit in 1962. Satellites were relatively cheap, especially when launched into geosynchronous orbits so they appeared stationary above the earth's surface. They promised much higher capacity than analog submarine cables. The American International Telecommunications Satellite Organization (Intelsat) was created in 1964 to manage the satellite system, although with geopolitical as well as commercial aims. Although the experimental satellites of the 1960s and 1970s had limited capacities Intelsat-V, launched in 1980, and -VI, launched in 1981, added 45,000 transatlantic telephone circuits between them.

Until the early 1990s Intelsat had as much of a telecommunications monopoly as Britain's Eastern Companies before World War I or Cable and Wireless in the late 1930s. As signals passed through Intelsat they were decrypted by the National Security Agency, giving America the same geopolitical advantages that Britain had previously accrued. Four events have reduced the commercial and geopolitical advantages of satellites. First, the 1986 failure of the American space shuttle, Challenger-7, threw the American launch program into disarray and resulted in most commercial satellite launches moving to the French Ariane program, although they would have likely done so anyway because the shuttle's cargo bay restricted satellite diameter. Second, the success of the noise- and delay-free fiber-optic cables after TAT-8 was laid in 1989, markedly reduced the utility of satellites for communications. Third, the technology of satellites is such that there are relatively few parking slots in geosynchronous orbit, they last only about ten years, and their stability is controlled by their diameter, which in turn is controlled by the launch vehicle. Fourth, rapidly improving encryption capabilities have made decryption impossible without acquiring the cryptography of other states through intelligence assets. Satellites have therefore had to find other uses: gathering remotely sensed data of human and natural activities on the earth's surface; for GPS location of mobile assets; and for digital television transmission. By the mid-1990s there were 30 satellites in geosynchronous orbit, most of which were launched aboard Arianes, 22 owned by Intelsat, 4 by Inmarsat of London, and 4 by PanAmSat of Greenwich, Connecticut.

Fiber-optics

In a talk given in 1969 Alec Reeves, the father of pulse code modulation (PCM), the digital encoding system universally used in telecommunications, forecast submarine fiber-optic cable using PCM within twenty years. The first, TAT-8, entered service in late 1988 carrying some 11,500 circuits and operating at 280 million bits per second with amplifiers about every 65 kilometers. Despite considerable problems developing an optically pure glass with low transmission losses and optical amplifiers that could be installed in the cables, fiber-optic technology has exploded since 1988. Capacity has skyrocketed and amplifier spacing

has lengthened. By the late twentieth century, 20 billion bits per second was possible in the laboratory. The target is a trillion bits per second through an individual fiber no bigger than a human hair, the equivalent of ten million simultaneous telephone calls.

Replacing analog with fiber-optic cables has returned emphasis to submarine transmission. Fiber-optic cables not only have very much higher capacity than analog ones, but also have the vanishingly low sampling errors important to PCM transmission. Laid over true terrestrial distances, they have none of the conversation interrupting transmission delays that occur in sending signals to a geosynchronous satellite and back to earth again, a fraction of a second even at the speed of light. Their low error rate makes them ideal for business data transmission, which is where most of their capacity is being used, although their vast overcapacity has led to extremely low cost for international telephony. Without them, the Internet would be impossible.

Cellular Telephones

The final great innovation in twentieth century communications has been the development of the mobile, or cell, phone. Using cellular technology this allows individuals to finally be free of the landlines that have connected them since the development of telegraphy in the 1840s. Cellular technology requires sophisticated computers to switch phones between cells and a network of fiber-optic landlines to link originating cells to the call's destination. The capacity of phones to carry data is still relatively limited, but visionaries propose cellular technology that would allow much more than mere telephony, up to and including individual mobile computing using "heads-up" interactive visual displays worn like eyeglasses.

See also **Electronic Communications; Fax Machine; Mobile (Cell) Telephones; Radio-Frequency Electronics; Satellites, Communications; Telephony**

PETER J. HUGILL

Further Reading

Aitken, H.G.J. *The Continuous Wave: Technology and American Radio, 1900–1932*. Princeton University Press, Princeton, 1985.

Beauchamp, K. *History of Telegraphy*. Institution of Electrical Engineers, London, 2001.

Burns, R.W. *The Life and Times of A D Blumlein*. Institution of Electrical Engineers, London, 2000.

Callick, E.B. *Metres to Microwaves: British Development of Active Components for Radar Systems, 1937–1944*. Peter Perigrinus, London, 1990.

Crowther, J. G. and Whiddington, R. *Science at War*. Philosophical Library, New York, 1948.

Douglas, S.J. *Inventing American Broadcasting 1899–1922*. Johns Hopkins University Press, Baltimore, 1987.

Edgerton, D. *England and the Aeroplane: An Essay on a Militant and Technological Nation*. Macmillan, London, 1991.

Guerlac, H.E. *Radar in World War II, Vol. 8: The History of Modern Physics, 1800–1950*. Tomash/American Institute of Physics, 1987.

Hecht, J. *City of Light. The Story of Fiber Optics*. Oxford University Press, New York, 1999.

Headrick, D.R. *The Invisible Weapon: Telecommunications and International Politics, 1851–1945*. Oxford University Press, New York, 1991.

Hook, S. *Wireless. From Marconi's Black-Box to the Audion*. MIT Press, Cambridge, MA, 2001.

Hugill, P.J. *Global Communications since 1844: Geopolitics and Technology*. Johns Hopkins University Press, Baltimore, 1999.

Jones, R.V. *The Wizard War: British Scientific Intelligence, 1939–1945*. Coward McCann, New York, 1978.

Mackinder, H.J. The geographical pivot of history. *Geographic. J.*, 23, 421–437, 1904.

O'Neill, E.E. *A History of Engineering and Science in the Bell System: Transmission Technology (1925–1975)*. AT&T Bell Laboratories, Indianapolis, IN, 1959.

Tilly, C. *Coercion, Capital, and European States, AD 990–1990*. Blackwell, Oxford, 1990.

Telephony, Automatic Systems

Fundamental to the expansion of telephone service was the widespread use of automatic telephone switches. Initial telephone service sought to connect every telephone directly with every other instrument in the same market. Soon manual switchboards were installed for more efficient operation. As more telephones entered service, switchboards became larger and more operators were required. Replacing often-slow manual operators, automatic devices were electromechanical for most of the first century of telephone usage, slowly being replaced by more efficient electronic (and eventually digital) systems.

Strowger Electromechanical Systems

The idea of an automatic telephone switch was first developed by Daniel and Thomas Connelly. Along with Thomas J. McTighe, they obtained the first patent on automatic or machine-switched telephony in 1879 (just a year after the world's first manual exchange had opened) and continued to develop their ideas, though their system was not commercially successful. More than 25 other

related patents followed from a variety of inventors, though no operating systems resulted.

The first practical automatic switch originated with the ideas of Almon B. Strowger, a Kansas City undertaker. Strowger was losing business to a competitor whose wife operated the local telephone manual exchange and could thus route business calls away from Strowger.

Seeking a way to eliminate the human operator, Strowger developed a stepped (or step-by-step, as the switch moved with every number entered) switch in 1886–1887, applied for his initial patent two years later while continuing to improve his system with the help of several collaborators, and formed a manufacturing company in 1890.

The first Strowger automatic step-by-step switch was installed in La Porte, Indiana (near Chicago) in 1892, serving about 75 subscribers and with a capacity of 100. The related telephone rotary dial instrument with its timed pulse was invented by a Strowger associate in 1896 (and introduced in a Milwaukee exchange) while an improved version appeared from the Ericsson firm that same year. By 1898 there were 22 automatic exchanges operating in the U.S. Other electromechanical systems included the semiautomatic system developed by Dane Sinclair in Britain (1883; installed in Glasgow, 1886), and the Lorimer system in Canada (1900), which although not successful, helped contribute to the later panel systems.

The Strowger system was "direct" in that the switch mechanically called up the telephone number as it received electrical pulses from the customer's dial, and is thus fast or slow depending on customer speed of use. Strowger switches were "progressive" in that the equipment does not "know" the next step (or number) until it is dialed. It utilized electromagnets to open and close mechanically linked connections. A metal selector or "wiper" moved both vertically and in a rotary horizontal fashion to reach the ten contact positions that made up each of the ten vertical levels. These were combined to create ever-larger switches. Strowger switches are often called "lost call" systems in that the constantly moving wiper seeks an unused line, feeding back a busy signal if one is not available.

Such systems were best suited to small or medium-sized markets (and indeed remained in use in some smaller towns until the end of the century). Large 10,000-line capacity Strowger switches had been developed by 1900 (the first was installed in New Bedford, MA). A decade later, more than 130 exchanges used Strowger equipment and served more than 200,000 customers. Use of such equipment was confined to independent (i.e., non-Bell System) companies. Versions of Strowger devices were also slowly taken up (and sometimes imported) by Canadian carriers and the major European powers (the first British public automatic exchange opened in Epsom, outside of London, in 1912, just as the British Post Office took over from private operators). Expansion of their use was often slow to develop due to a readily available and inexpensive female labor pool. In the U.S., however, World War I labor demands helped to push wider usage.

By the 1920s only about 15 percent of the world's telephones used automatic switches of some kind, although most major American cities were served by such switches by 1939. Half of German telephones were served by automatic switches by 1930; Britain achieved that mark six years later.

Rotary, Panel, and Crossbar Electromechanical Systems

The short-lived Lorimer switch gave rise to two systems, both of which originated with Western Electric. The "rotary" automatic system utilized a cylindrical wiper that could connect with 30 different positions for a total of 300 lines per switch. Though AT&T decided not to use it, the rotary switch was widely adopted after World War I in Europe and other countries and successive developments remained in use from about 1910 into the 1970s.

"Panel"-style automatic switches were first experimented with by Bell Systems in 1912 and after considerable research, an improved variation was introduced in 1921in Omaha, Nebraska. These were more complex indirect-control devices (speed of dialing had no impact on how the switch worked) that used moveable vertical rods to switch between 500 lines (five times greater than contemporary Strowger gear) arranged across a wide panel. They were expensive to maintain and were used only by the Bell System (in 26 large cities by the late 1920s), slowly giving way to the crossbar system after the 1950s.

Based in part on the panel system, and on Swedish inventions early in the century as well as the crossbar switch of 1913 developed by John N. Reynolds of AT&T's Western Electric, the first experimental "crossbar" exchange opened in Sweden in 1926. Convinced of its merits, AT&T introduced an improved version in early 1938, its No. 1 crossbar switching system, designed for large urban centers and first installed in Brooklyn, NY.

The crossbar switch used common control (call handling was totally separated from interconnection with the network), with switches making only small motions (allowing greater speed), thus suffering less vibration than the Strowger, and allowing greater flexibility. It consisted of relays, complex link-trunking, and allowed alternative routing of calls. The No. 5 crossbar switch eventually achieved a capacity of 35,000 lines, but could easily be adapted for smaller exchanges.

Their higher costs (compared to existing panel switches) delayed installation of improved crossbar systems, but eventually more efficient crossbars (the No. 4 and No. 5, introduced in the Philadelphia area in 1943 and 1948, respectively) were developed. In the 1950s similar crossbar systems also began to take hold in Europe and Japan. By this time, more than three quarters of the world's telephones were served by automatic switches of one kind or another.

Toll or long-distance automatic switching developed more slowly, however, and required development of universal telephone numbering plans (including the use of area codes) for different parts of the world. Direct distance dialing was introduced in the U.S. in 1951, using the capable No. 5 crossbar switch. While the first international automatic service was introduced between Brussels and Paris in mid-1956, initial customer direct-dialed international calls (New York to Britain) became possible only in 1970. Crossbar switches manufactured by several firms were soon the most widely used automatic switches in the world, and continued to appear in improved versions into the 1970s. Many remained in use beyond the end of the century, proving them to be highly reliable and efficient over time.

Electronic Systems

Based on postwar computer and transistor developments, experiments with fully electronic switches began at Bell Laboratories in the early 1950s. Offering the potential of much higher speeds and lower power requirements (and thus smaller equipment designs), the possibility of such switches became more viable with innovations in solid-state electronics and the development of computer stored-program control (SPC) after 1960. After 500 million dollars of research and experimentation expenditure and considerable delay, AT&T's No. 1 ESS (electronic switching system) was first installed in a New Jersey suburb of New York in 1965.

With ESS, electronic scanning of telephone lines replaced use of mechanical relays and switches. ESS systems utilized complex computer programming and parallel processing both to determine system faults and to provide service redundancy and thus reduce downtime. Reliability was said to allow no more than two hours of downtime over a 40-year period. Within a decade, more than 500 offices were served by ESS installations. The No. 2 ESS for smaller markets followed, the first of which entered service in 1970. By the 1980s more than 700 offices were served by No. 2 ESS equipment.

The first time-division all-digital (or second-generation) electronic switching system was the No. 4 ESS, introduced in Chicago and other major American cities in 1976, which could handle nearly 54,000 two-way voice circuits (and up to 500,000 call attempts) per hour. The No. 5 ESS was delayed by developmental problems and was first introduced in 1982 in Seneca, Illinois. It was also the first switch that AT&T had to sell in a competitive market, because after 1984 its operating companies were independent. After 1996, Lucent Technologies continued development and marketing of the system by which time an improved No. 5 ESS could handle up to a million call attempts per hour.

See also **Packet Switching; Telephony, Digital**

CHRISTOPHER H. STERLING

Further Reading

Aitken, W. *Automatic Telephone Systems*, 3 vols. Benn Brothers, London 1921–1924.

Baldwin, F.G.C. Automatic or machine switching systems, 1879–1922, in *The History of The Telephone in the United Kingdom*. Chapman & Hall, London, 1925.

Chapuis, R.J. Manual and electromechanical switching (1878–1960s), in *100 Years of Telephone Switching (1878–1978)*. North Holland, Amsterdam, 1982.

Chapuis, R.J. and Amos Jr., E.J. Electronics, computers and telephone switching: a book of technological history, in *100 Years of Telephone Switching, Part 2: 1960–1985*. North Holland, Amsterdam, 1990.

Hanscom, C.D. *Dates in American Telephone Technology*. Bell Telephone Laboratories, New York, 1961.

Joel Jr., A.E. and Schindler Jr., G.E., Eds. *A History of Engineering and Science in the Bell System: Switching Technology (1925–1975)*. Bell Telephone Laboratories, New York, 1982.

Moudon, R. *The Strowger Automatic Telephone Exchange*. Spon & Chamberlain, New York, 1919.

Noll, A.M. *Introduction to Telephones and Telephone Systems*, 3rd edn. Artech House, Norwood, MA, 1998.

Ryan, J.S. Signalling and switching as we enter the second century. *Telecomm. J.*, 43, 206–219, 1976.

Smith, A.B. and Campbell, W.L. *Automatic Telephony: A Comprehensive Treatise on Automatic and Semi-Automatic Systems*, 2nd edn. McGraw-Hill, New York, 1921.

Switching and signalling. *Post Office Electr. Eng. J.*, 74, 187–222, 1981.
Various. The development of the automatic telephone, 1889 through 1918. *Telecom History*, 2, 1–112, 1995.

Telephony, Digital

Digital telephony is the digital transmission of voice over a communications network. This technology entails digital encoding of voice and digital transmission with regenerating repeaters, switches, and handsets. Using digitally enabled multiplexing, voice is integrated with computer data and digitized images, video, and other information in current telephone and computer networks.

The Bell Telephone Company was formed by Alexander Graham Bell in 1878 as the first U.S. telephone system. The public switched telephone network (PSTN) evolved into an analog, circuit-switched, internationally connected telephone system. In analog transmission, signals proportional to voice input are propagated throughout the length of the channel. In circuit-switched networks, a connection is established and bandwidth is dedicated for the entire connection. Although human voice extends over a 20-kilohertz range, voice grade lines have been limited to a 4 kilohertz band in order to decrease resource use.

In 1939, pulse code modulation was developed by Alex H. Reeves as a method for encoding an analog signal such as voice with digital values. A coder–decoder (codec) samples the analog, 4-kilohertz voice band at 8000 times per second, which is sufficient (according to Nyquist's theorem, 1924), to capture the entire information in that band. Each sample is assigned one of 128 values (North America) or 256 values (Europe) and encoded in seven or eight bits, requiring a data rate of 64 kilobits per second (Kbps), including control information. Approximating continuous analog values by discrete levels, called quantizing, results in some information loss. For high-quality voice in audio compact disks, a 22-kilohertz band is sampled at 44,100 samples per second, with 16-bit samples encoded over 65 kilohertz levels, requiring a 705.6 Kbps (double for stereo) data rate without compression if this data is transmitted. Differential compression schemes, such as differential pulse code modulation and delta modulation, as well as predictive schemes, such as adaptive differential pulse code modulation, code excited linear prediction and multi-pulse excited linear predictive coding, allow lower transmission rates, but at decreased signal quality.

Once voice was digitized, it became feasible to combine several input lines into a round-robin output stream. In 1957, Bell Laboratories developed the T-1 carrier system, which was implemented by 1962 in a Chicago area exchange. This digital timed-based multiplexing system takes turns allocating 8 bits of each of 24 input lines to the output stream. The total output transmission rate of 1.544 megabits per second (Mbps) includes 8 Kbps for clock synchronization. In Europe, an equivalent E-1 carrier multiplexes 32 input lines, each at 64 Kbps, for a 2.048 Mbps output. T-1 lines are combined to form T-2, T-3 and T-4 lines for higher transmission rates; E-1 lines are similarly bundled together.

"T" carrier systems were soon directly connected to digital switches. AT&T's electronic switching systems were installed in private computerized branch exchanges (PCBX) by 1963 and in public network telephony by 1965. Manual functions were computerized in large digital switches, which were able to combine the functions of several analog switches. Very large-scale integration (VSLI), spearheaded by Intel in 1971 for the manufacture of computers, made these systems cost effective.

The PSTN's backbone system was mainly based on analog radio trunks in the 1970s, with some reliance on coaxial cable. A major impetus for the transition from analog to digital transmission was the development of optical fibers. Prototypes of optical fibers that were suitable for transmission were developed at Corning Glass Works by 1970. By 1977, General Telephone and Electric (GTE) had installed optical-fiber commercial systems. Fiber's high bandwidth, low attenuation, small diameter, and immunity to electromagnetic interference spearheaded the replacement of analog trunks with digital fiber trunks.

By the 1980s, AT&T had replaced its entire analog backbone, implementing Integrated Digital Networks (IDN), the digital transmission of voice and data throughout its backbone network. By the 1990s, Integrated Services Digital Network (ISDN) extended digital telephony to the local loop, providing digital services directly to customers in many locations. Many new functions were enabled, including automated call tracing, automatic call back, call waiting, and call forwarding. Motivated by the high bandwidth demands of the Internet, other end-to-end digital access technologies were developed in the 1990s, notably digital subscriber lines (xDSL) and cable modems.

Digital telephony has evolved in wireless systems as well. As mobile telephony was being

developed during the 1940s to 1960s, only a limited number of channels (initially 11 in the 40-megahertz band) were available. Blocking and high costs were common. Each (very heavy) handset searched the frequency spectrum, looking for an idle channel. In 1981, the first modern cellular (analog) system, the Nordic Mobile Telephone System, went into commercial operation in Sweden. Advanced Mobile Phone System (AMPS), a cellular analog system developed at Bell Labs, became commercially available by 1983. More recently, Digital Advanced Mobile Phone System (D-AMPS) has become available in the U.S., as well as Global System for Mobile (GSM) communications, a digital cellular system that was developed in Europe by 1982, and other systems. A digital wireless metropolitan area network may be used for the local loop, connecting customers to the PSTN, following the standards published by Institution of Electrical and Electronics Engineers (IEEE 802.16, in 2002. Satellite systems, such as Iridium (1998), Globalstar (2000), and Teledesic (under development), on the other hand, promise end-to-end digital telephony.

Since telephony software became commercially available in 1995, Internet telephony (VoIP) has connected the Internet to the PSTN using speakers and microphones on personal computers. The Internet, which is a connection of computer networks all over the world, is based on packet-switching technology. Packet switching (see separate entry) breaks the data stream into small units and adds control fields to each unit, for functions such as addressing and error correction. Packet switching allocates bandwidth on a demand basis, making it difficult to guarantee the timely and predictable service that is necessary for good quality voice reception. PacketCable, designed for Internet protocol (IP) technology over cable modems, includes quality of service options for better reception. Several application programming interfaces have been defined to connect computer systems to telephone services, such as TAPI, TSAPI, and JTAPI.

Digital telephony has rapidly replaced analog telephony due to many factors: VSLI technology provides lower costs; discrete values sampled by receivers and regenerating repeaters enable recovery from most transmission noise; digital methods of multiplexing, such as time division multiplexing (TDM) and code division multiple access (CDMA), afford higher throughput. As VoIP, the PSTN, and satellite telephony continue to develop, will packet switching and circuit switching continue to coexist and will a single technology dominate in digital telephony?

See also **Mobile (Cell) Telephones; Packet Switching; Optoelectronics, Dense Wavelength Division Multiplexing**

TRUDY LEVINE

Further Reading

Bellamy, J. *Digital Telephony* 3rd edn. Wiley, New York, 2000.

Griffiths, J.M. *ISDN Explained*. Wiley, New York, 1990.

Tanenbaum, A. *Computer Networks*, 4th edn. Prentice Hall, Upper Saddle River, NJ, 2003.

Wright, D.J. *Voice Over Packet Networks*. Wiley, New York, 2001.

Telephony, Long Distance

Originally a means of communication suited only for small or medium distances, from about 1920, cable and wire telephony—soon to be supplemented by radio telephony—began to conquer the longer distances as well. Advances in telephone telephony, such as the post-World War II transoceanic cables, became a key factor in the trend towards the global information society. Much of the progress was based in new scientific knowledge or transformation of such knowledge into technological methods and artifacts.

Shortly after Alexander Graham Bell's invention of the telephone in 1876, it became clear that it was technically difficult to transmit speech signals over long distances. By using gutta percha-insulated cables and thick copper wires the distance was increased, but this traditional technology did not allow economically feasible lines much longer than 1000 kilometers. The first major innovation was based on the theory of telephone current transmission independently proposed in 1887 by Oliver Heaviside in England and Aimé Vaschy in France. According to this theory, the attenuation would be minimized, hence the speaking distance increased, if the line was "loaded" with self-inductance (the counter-electromotive force arising from the rapid variation of currents). Only around 1902 was the theoretical insight turned into a practical method of how to add self-inductance to wires and cables. In the U.S., George Campbell and Michael Pupin devised methods for discrete loading by inserting induction coils in the line. The Danish engineer Carl Krarup invented an alternative method of continuous loading by winding the copper conductors with thin wires of iron. The loading method was highly successful and soon used for almost all

submarine cables and long landlines. Loading technology increased the speaking distance, but not dramatically. The big step forward in long-distance telephony proved to be the introduction of the vacuum tube as an amplifier or "repeater" of the weak speech currents. In 1915, Harold Arnold and his staff at American Telephone and Telegraph Company (AT&T) had developed a tube repeater that, in combination with coil loading, made it possible to telephone from New York to San Francisco. After the end of World War I, tube-amplified lines formed the basis of the first long-distance telephone networks in Europe and the U.S. By the late 1920s, almost all European countries were connected by telephone lines.

In no area of telephony was the power of the vacuum-tube repeater more impressive than in long-distance submarine telephony. Loading technology increased the efficiency and range of submarine telephone cables, and in 1928 AT&T engineers made a proposal of designing a transatlantic cable based on continuous loading. Nothing came of the plan, however, and with the development during World War II of submerged tube repeaters it was realized that this was the method to be used in a future transatlantic cable. The first cable crossing the Atlantic, named TAT-1, was completed in 1956 as a joint project of AT&T, the British Post Office, and the Canadian Overseas Telecommunication Corporation. The heart of the 4,244-kilometer cable was the 51 submerged tube repeater units. Subsequent ocean telephone cables were even longer. In 1963 a distance record was achieved with the laying of a transpacific cable between Canada and Australia over a distance of nearly 11,000 kilometers. Whereas the first generation of transoceanic cables used vacuum tubes in the repeater units, from the mid-1960s tubes were increasingly replaced by transistors.

TAT-1 made crucial use of another important innovation in telephone technology—the frequency-division-multiplex (FDM) coaxial system, an advanced form of carrier-wave telephony in which many messages can be sent simultaneously through a single wire. In 1918, AT&T inaugurated the first commercial carrier system, and in 1931 the company had developed the first submarine coaxial carrier cable for use between Havana and Key West. The TAT-1 coaxial cable was insulated with polyethylene and included 36 voice channels, a number that would increase drastically in later transatlantic cables. For example, the transistorized TAT-6 cable of 1976 had a capacity of 4,000 channels. The largest cable of the traditional coaxial type was the 15,000-kilometer-long ANZCAN cable between Canada and Australia that started service in 1984. Investments amounted to $500 million and each of the 1,124 transistorized repeaters cost nearly $100 000. With the invention of the laser in 1960 and the development of low-attenuation glass fibers about 1970, fiber-optical telephone communications became a possibility. The new cables were immediately used for submarine cables, as in 1988 when TAT-8 was opened for service between England and North America. This project, a collaboration between AT&T, British Telecom, and France Telecom, had a capacity that doubled the existing cable capacity across the Atlantic. In 1991 it was followed by a fiber-optic link between the U.S. and Japan.

With the development of vacuum tubes, wireless telephony became a possibility for communication over long distances. Experimental radio telephony across the Atlantic was first achieved in 1915 by AT&T, and six years later the same company developed a hybrid communication system that combined inductively loaded cables and wires with radio links. The breakthrough of wireless telephony occurred a few years later, with the introduction of short-wave radio transmission. In 1927 the first commercial transatlantic radiotelephone service started between London and New York, and within a few years intercontinental telephony over the combined radio and cable network was a reality. At that time, truly intercontinental telephony, defined as circuits joining different continents over a distance of more than 1,000 kilometers, was limited to radio links. For example, the length of the 1930 London to Sydney link was 17,000 kilometers, and in 1934 AT&T performed the first round-the-world telephone conversation through combined wired and wireless circuits.

During the last three decades of the century, an increasing part of intercontinental telephony was transmitted wirelessly through satellites rather than through cables. Many experts believe that future long-distance communications will rely on a mixture of optical-fiber cables and satellites, with a major part of the telephone traffic going through the satellite service.

See also **Mobile (Cell) Telephones; Radio Transmitters; Satellites, Communications; Telecommunications; Telephony, Digital**

HELGE KRAGH

Further Reading

Bray, J. *The Communications Miracle: The Telecommunication Pioneers from Morse to the*

Information Superhighways. Plenum Press, New York, 1995.

Brittain, J.E. The introduction of the loading coil: George A. Campbell and Michael I. Pupin. *Technol. Cult.*, 11, 36–57, 1970.

Fagen, M.D. *A History of Engineering and Science in the Bell System: The Early Years, 1875–1925*. Bell Telephone Laboratories, Whippany, NJ, 1975.

Faltas, S. The invention of fibre-optic communications. *Hist. Technol.*, 5, 31–49, 1988.

Headrick, D.R. Shortwave radio and its impact on international telecommunications between the wars. *Hist. Technol*, 11, 21–32, 1994.

Kragh, H The Krarup cable: invention and early development. *Technol. and Cult.*, 35, 129–157, 1994.

Kragh, H. History and prehistory of the first transatlantic telephone cable. *Polhem*, 13, 246–271, 1995.

Solymar, L. *Getting the Message: A History of Communications*. Oxford University Press, Oxford, 1999.

Wasserman, N.H. *From Invention to Innovation: Long-Distance Telephone Transmission at the Turn of the Century*. Johns Hopkins University Press, Baltimore, 1985.

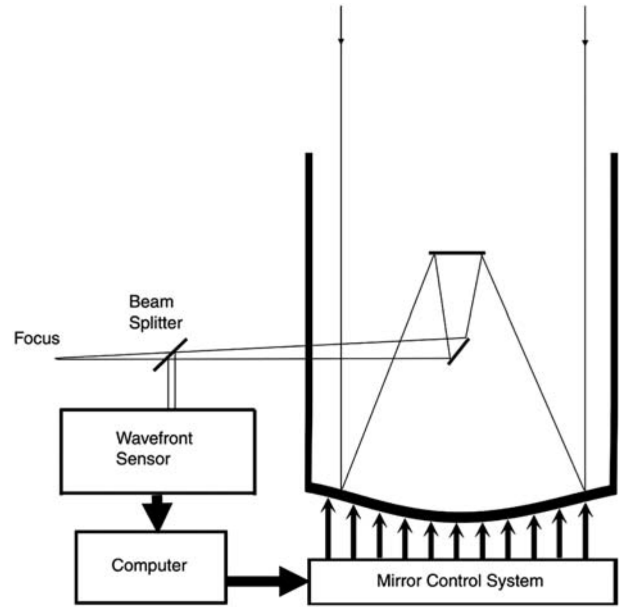


Figure 1. Active optics system.

Telescopes, Computer-Controlled Mirrors

The first half of the twentieth century saw a phenomenal increase in the size of astronomical telescopes, culminating with the 5-meter Hale instrument on Mount Palomar. However by mid-century it appeared that the ultimate size of earth-bound telescopes was inherently limited. Despite the use of new materials and advanced engineering, increasing the size of telescope mirrors seemed to make them inherently more sensitive to flexure and vibration. Moreover, the resolving power of earth-bound telescopes seemed to be ultimately limited by the distorting effects of atmospheric turbulence. One of the primary justifications for funding the Hubble space telescope was to avoid these difficulties.

In the second half of the twentieth century astronomers explored new technological approaches to solving these problems, many of which concentrated on developing increasingly sophisticated sensing devices. But there was also an impetus to rethink the telescope from its traditionally passive role and to give it some capacity to react to changes in viewing conditions. Eventually two separate, but complementary, systems emerged: active optics and adaptive optics.

Active optics involves changing the shape of the telescope's primary (and in some cases, secondary) mirror to optimize the instruments' resolving power (see Figure 1). Active optic systems analyze the image of a reference object (typically a bright star located near the object being studied) to generate a series of changes in the telescope's optical system, optimizing the image of the object being viewed. These changes may be performed relatively slowly and can correct for a variety of problems, including low-frequency wind vibrations, uneven thermal expansion, gravitational distortion and even the instrument's permanent figuring errors. Instruments with active optical systems can be made significantly lighter than conventional telescopes and are well suited to use on mirrors made up of multiple segments.

Adaptive optics involves changing the shape of a much smaller mirror, typically positioned within the telescope's optical train, with the goal of correcting wave-front distortions that occur as light passes through the earth's atmosphere (see Figure 2). Adaptive optics uses a guide-star feedback loop similar to that used in active optics but with much smaller corrections that are ideally made in the space of a few milliseconds.

Although suggested as early as 1953, practical implementation of these systems required the development of a variety of supporting devices, including wave-front sensors, deformable mirrors, actuators, computers, and computer software. While active optics is now considered a mature technology, the routine employment of adaptive optics to the entire visual spectrum is still under development. Significant correction of the atmospheric distortion of infrared light has been

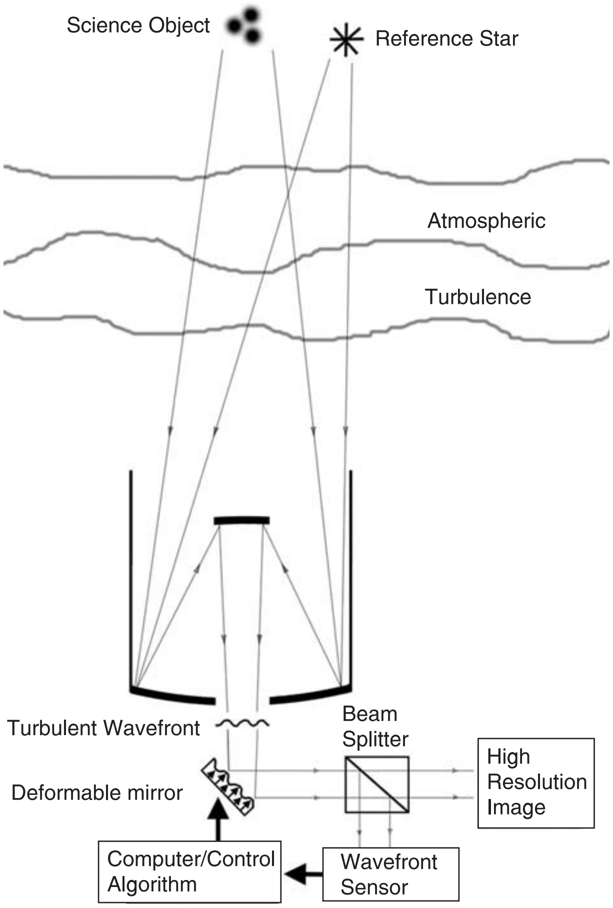


Figure 2. Adaptive optics system.

successfully achieved, but correction of distortion to higher frequencies remains a challenge.

Another important challenge has involved finding suitable reference stars for image and wave-front correction. The ideal reference is a bright star located fairly close to the object being studied but far enough away from the telescope's optical axis that its light is not part of the observation. The source needs to be bright enough to overcome noise in the wave-front detectors and it needs to be close enough to the science object to have roughly the same light-path through the atmosphere. Finding a suitable natural guide star (NGS) for every observation has proved to be a problem and the possibility of generating artificial reference stars was an early area of investigation.

This work received an unexpected boost from the U.S.'s Strategic Defense Initiative (SDI) in the 1980s. SDI planned to use a form of adaptive optics to focus a destructive laser beam on incoming missiles. To permit rapid response, the system used an artificial reference star created by shining a laser on a naturally occurring layer of sodium atoms approximately 90 kilometers above the earth's surface. When the SDI program was cancelled in 1991 much of this technology was turned over to astronomers. The use of laser guide stars (LGS) in astronomy dates from this time.

While the use of LGS appears to solve the reference star problem, the relative nearness of these objects introduces the problem of conical anisoplanatism. This refers to the fact that light from the LGS arrives at the telescope in a path that is cone shaped. This can cause the LGS wave-front distortion to differ significantly from that of astronomical object being studied. LGS systems can also be expensive to install and operate. A number of alternatives have been proposed, including the use of multiple guide stars (either natural or artificial) to produce a tomographic AO correction for entire sections of the sky.

Progress in both active and adaptive optical systems continues to accelerate and it is clear that future earth-bound astronomical telescopes will depend heavily on them. While complete correction of observing errors is not possible, even partial correction can produce dramatically improved images. The image quality of modern astronomical telescopes is commonly expressed in terms of the strehl ratio, which is defined as the ratio of the peak intensity of an image divided by the peak intensity of a diffraction-limited (ideal) image. A perfect strehl ratio is 1.0; for reference, the Hubble space telescope (HST) is estimated to have a strehl ratio of 0.97.

For earth-bound telescopes, the highest achievable strehl may be around 0.8 and day-to-day observations may fall short of that goal. However it has already been demonstrated that, using adaptive optics, a 10-meter surface telescope can produce infrared images that surpass those produced by Hubble. This is because of the much greater light-gathering area of the surface telescope. Moreover, the cost of building and operating large earth-based instruments can be significantly lower than launching telescopes into space. With the size of launchable space telescopes currently limited, attention is being increasingly directed towards the construction of large AO earth based observatories. The next generation of visual telescopes may be in the 20 to 30 meter range and instruments as large as 100 meters have been proposed.

See also **Telescopes, Ground; Telescopes, Space**

STEVEN TURNER

Further Reading

Babcock, H. The possibility of compensating astronomical seeing. *Public. Astronom. Soc. Pacific*, 65, 229–236, 1953.

Babcock, H. Adaptive optics revisited. *Science*, 249, 253–257, 1990.

Busher, D. Love, G. and Myers, R. *Adaptive Optics*. Wiley, New York, 2003.

Le Louarn, M., Hubin, N., Sarazin, M. and Tokovinin, A. New challenges for adaptive optics: extremely large telescopes. *Mon. Notices R. Astronom. Soc.*, 317, 535–546, 2000.

Marshall, E. Researchers sift the ashes of SDI. *Science*, 263, 620–623, 1994.

Ragazzoni, R., Marchetti, E. and Valente, G. Adaptive-optics corrections available for the whole sky. *Nature*, 403, 54–56, 2000.

Roddier, F., Ed. *Adaptive Optics in Astronomy*. Cambridge University Press, Cambridge, 1999.

Taroy, M., Ed. *Adaptive Optics: Current Abstracts With Indexes*. Nova Science, New York, 2002.

Useful Websites

Center for Astronomical Adaptive Optics, Steward Observatory, University of Arizona: http://caao.as.arizona.edu/publications/. Provides an extensive bibliography.

Telescopes, Ground

A telescope collects radiation from distant sources and converges it onto a detector such as the human eye or photographic plate. Historically astronomers have desired bigger telescopes that collect more light and enable scientists to observe fainter, more distant objects. Throughout the twentieth century, scientists and engineers have proposed novel designs for telescopes with increased light-collecting power. Choosing and implementing these has required a combination of community support, technological capability, and financial resources.

In 1900 astronomers still used two basic types of telescopes: refracting and reflecting. Refracting telescopes, which use a system of lenses to form an image, were precision instruments preferred by many professional astronomers for tasks such as measuring the positions of stars. Reflecting telescopes use glass mirrors coated with a thin layer of reflective metal to collect and focus light. Reflectors did not suffer from intrinsic optical defects that limited the usefulness of refracting telescopes. Somewhat easier to use and enclosed in smaller and less expensive domes, reflecting telescopes became the central instrument of the modern observatory as they were better suited for astrophysical studies of celestial objects via spectroscopy and photography.

The person perhaps most responsible for the development of large telescopes in the first half of the twentieth century was George Ellery Hale. Scion of a prominent Chicago family, Hale moved to California in 1903 and established the Mount Wilson Observatory overlooking Pasadena. Generous funding from the Carnegie Institution of Washington enabled Hale to commission specially designed telescopes to observe the sun. Hale later obtained a 60-inch (152-centimeter) glass disk which became the heart of the first large reflecting telescope at Mount Wilson. Even before workmen finished the 60-inch in 1908, Hale already had plans for a bigger telescope. Built with funding from a wealthy merchant and the Carnegie Institution, the 100-inch (254-centimeter) telescope entered service in 1919. It collected three times more light than the 60-inch and played a major role in reshaping cosmological theory. Astronomers like Edwin P. Hubble, for example, used it to establish the nature and distance of nebulae and provide evidence the universe was expanding.

In 1928 Hale began campaigning for a 200-inch (508-centimeter) telescope. He secured $6,000,000 from the Rockefeller Foundation and began plans for the world's largest telescope to be jointly run by Mount Wilson and the California Institute of Technology. The telescope was built on Palomar Mountain in southern California. Building the 200-inch telescope, especially its 20-ton mirror, was a Herculean effort and it was not dedicated until 1948, ten years after Hale's death.

The 200-inch defined what a large telescope should look like for the next three decades. Many slightly smaller telescopes; for example, the Kitt Peak 4-meter telescope and the Anglo–Australian Telescope, were based on its design. They featured tremendous dome enclosures, primary mirrors made of a single massive piece of glass, an immense horseshoe-shaped bearing, and traditional equatorial mounts. It was not until the Soviet Union completed its 6-meter *Bolshoi Telescop Azimutal'ny* in 1976 that astronomers had a telescope larger than the 200-inch. Despite the BTA's innovative altitude-azimuth mount, its capabilities were limited for years by a poorly performing mirror while Cold War politics limited its usefulness to Western astronomers. An altitude-azimuth mount allows the telescope to pivot simultaneously on up–down around one axis (in altitude) and horizontally (in azimuth) around the other. This was made possible by the availability of computer control in the 1970s.

All of the telescopes described thus far gather visible light. Telescopes can also collect other types of radiation. For example, astronomers and engineers built telescopes such as the United Kingdom Infrared Telescope (finished in 1979 on Mauna Kea in Hawaii) that were optimized for infrared observing. Telescopes can also capture gamma rays for high-energy astrophysics. When gamma rays pass through the earth's atmosphere they generate Cherenkov radiation, which specially designed telescopes can detect on the ground as flashes of blue light. Another variation are the "neutrino telescopes" located underground to detect elusive subatomic particles produced by reactions deep inside the sun.

While the 200-inch remained the largest optical telescope in the Western world for over 40 years, postwar astronomers devised many ways to collect more photons. One solution was electronic detectors such as image tubes and photomultipliers. These electronic devices were many more times efficient than photographic techniques, effectively giving scientists greater light-collecting power by recording more photons collected by a telescope. Electronic technologies refined after 1950 encouraged astronomers to build more sophisticated light detectors instead of bigger telescopes, which were far more expensive. Astronomers benefited from this phase of technological development well into the 1970s with the introduction of even more sensitive CCD (charge-coupled devices) detectors.

In 1980 a new telescope was introduced that broke from the traditional design established by the 200-inch. The Multiple Mirror Telescope (on Mount Hopkins in southern Arizona) featured six 1.8-meter mirrors on a common mount. These combined light collected at a common focus, making the MMT equivalent to a 4.5-meter telescope. The MMT's optical system—now converted to a single mirror—was innovative as was its use of a compact altitude-azimuth mount, extensive computer control, and inexpensive and efficient shed-like enclosure that rotated with the telescope. With the MMT astronomers showed they could collect more light at a reasonable cost.

From the 1980s astronomers and engineers developed several techniques for making large, lightweight, yet relatively inexpensive mirrors around which new telescopes were built. Jerry Nelson and his colleagues at the University of California based their design for the two 10-meter Keck telescopes on Mauna Kea around a mirror made of 36 smaller glass segments; scientists at the University of Arizona led by Roger Angel developed an innovative way to spin-cast large mirrors; and commercial firms such as Corning introduced thin "meniscus" mirrors. Engineers have used each of these solutions in a number of large telescope projects. Another potentially important development in the late 1980s was adaptive optics. Derived from military systems built for missile defense and satellite tracking, adaptive optics uses a complex optoelectonic system which can be adjusted many times a second. This offers astronomers the means to remove distortion caused by atmospheric turbulence and produce sharper pictures.

In 1980 the light collecting area of optical telescopes worldwide was about 150 square meters; in 2002 it was over 900. Astronomers use new telescopes differently than they did when the venerable 200-inch entered service. In Edwin Hubble's time astronomy was a mostly a solitary pursuit as the scientist worked alone at the telescope collecting data on photographic plates. Fifty years later ground-based optical astronomers sit in a warm, well-lit control room and monitor their data on computers. Astronomers can also operate some telescopes remotely from thousands of kilometers away while electronic data archiving may enable scientists to access far greater amounts of information. The increased size and sophistication of new telescopes have helped alter what it means to be an astronomer. As astronomers discuss plans for even larger telescopes—30 to 100 meters in size—and court potential patrons, the sociological effects of new and more complex telescope technology remains to be seen.

See also **Telescopes, Radio; Telescopes, Computer-Controlled Mirrors**

W. PATRICK MCCRAY

Further Reading

Florence, R. *The Perfect Machine: Building the Palomar Telescope*. Harper Perennial, New York, 1994.
King, H. *The History of the Telescope*. Sky Publishing Corporation, Cambridge, 1955.
Learner, R. *Astronomy Through the Telescope*. Van Nostrand Reinhold, New York, 1981.
Maran, S. A new generation of giant eyes gets ready to probe the universe. *Smithsonian*, 1987, 41–53.
Mountain, M. and Gillet, F. The revolution in telescope aperture. *Nature* 395, A23–29, 1998.
Osterbrock, D. *Pauper and Prince: Ritchey, Hale, and Big American Telescopes*. University of Arizona Press, Tucson, 1993.
Overbye, D. *Lonely Hearts of the Cosmos: The Story of the Scientific Quest for the Secrets of the Universe*. Little, Brown & Company, New York, 1991.
Smith, R. Engines of discovery: scientific instruments and the history of astronomy and planetary science in the

United States in the 20th century. *J. Hist. Astron.*, 28, 49–77, 1997.
van Helden, A. Building large telescopes, 1900–1950, in *Astrophysics and 20th Century Astronomy to 1950: Part A*. Gingerich, O., Ed. Cambridge University Press, Cambridge, 1984, pp. 134–152.

Telescopes, Radio

Radio telescopes are instruments for the study of cosmic radio emissions. In a strict sense, a radio telescope is any apparatus designed to detect radio waves—waves with wavelengths in the range between 1 millimeter and 30 meters—of extraterrestrial origin. The first designs of antennae and receptors to detect these waves were made in the late nineteenth century by scientists such as Americans Thomas A. Edison and Arthur E. Kennelly, the British Oliver J. Lodge, the Germans Johannes Wilsing and Julius Scheiner, and the French Charles Nordman. The purpose of these designs was the study of the sun as a source of radio waves, but no results were obtained. It was in 1942 when the British J. Stanley Hey for the first time successfully detected radio waves from the sun while studying the interference of British radar during World War II. Radio astronomy became a recognized part of astronomy during the 1950s, although successful experiments of that kind had been done as early as in 1932. This was the year in which the American radio engineer Karl G. Jansky, studying interference in transatlantic communications for the Bell Telephone Laboratories, identified short-wave noise resulting not only from thunderstorms, but also from an extraterrestrial source. As we now know, the signal he detected was synchrotron emission associated with energetic electrons accelerated in the magnetic field of the Milky Way galaxy. However, astronomers did not follow up Jansky's results. Other work, such as that of the American Grote Reber, who mapped the sky during the 1940s using contours of high and equal radio-wave intensity, required the design of new parabolic dish antennae suitable for detection of higher frequencies. The 10-meter parabolic reflector dish that Reber built in 1937 in his backyard to seek cosmic radio emissions is generally considered the first radio telescope used for astronomical research.

It was not until the end of World War II that radio astronomy received firm support. The equipment that was now freely available for use in this new field of astronomy and the knowledge acquired in the area of radar and radio communications made this possible: the great advances in microwave technology that had led to radar became available to astronomers. In fact, the instruments that were used during that period were not very different from the ones used nowadays, at least as far as the basic structure is concerned. In principle, a radio system consists of two basic elements: the large radio antenna and the radiometer or radio receiver. The antenna controls the direction of observation and then collects the radiation and converts it into an electric (alternating current) signal. This part of the radio system is very sensitive to polarization of the incident radiation, which is why crossed antennae are usually used to record the whole signal of the radio source. Antennae must be carefully designed in order to minimize the effects produced by different responses at various angles. On the other hand, it is the receiver that picks out from the signal a particular frequency and bandwidth to work with, which is subsequently processed and recorded. One of the problems radio astronomers have to deal with has been that for longer wavelength, larger apertures of telescopes are needed to achieve good resolution. Radio telescopes are limited by the largest practicable aperture, as well as by the diffraction produced by the receiver itself. A solution adopted in the second half of the 1940s emerged from working on solar eclipses. However, still better solutions were required in order to work regularly with radio telescopes without obtaining distorted images.

While the sensitivity of both radio and optical telescopes depends on the total surface area of the collector of radiation, the angular resolution of a radio telescope essentially depends on the largest distance between two points on the antennae system, which pick up signals to be processed simultaneously in the receiver. This distance is referred to as the aperture width. The larger the aperture width, the better the resolution. The aperture width is essentially equal to the diameter in what are called "filled aperture radio telescopes." These are frequently single-dish and their design is similar to optical telescopes. Another way of increasing the effective aperture width is using an array of radio telescopes. In other words, single telescopes at different places pick up the signal to be processed simultaneously, increasing the angular resolution. Martin Ryle and D. Vonberg at Cambridge, U.K., first employed this technique, known as radiointerferometry, in the 1940s. The particular technique used, still known as earth rotation aperture synthesis, was a two antennae radio interferometer with a maximum spacing of about 0.5 kilometers. Interferometric techniques at radio wavelength were developed

during the 1950s and the 1960s. The highest resolution results were obtained with arrays composed of big single radio telescopes separated by very large distances. In these cases, cables, of course, do not link the components. Two typical examples of how these radio telescopes work are the British MERLIN (multi-element radio-linked interferometer network) begun in 1976 and extended in the 1990s, which works with radio links, and the VLA (very large array) in New Mexico dedicated in 1980, which uses a technique called VLBI (very long-baseline interferometry) where signals are recorded separately and combined later by computer. The latter currently represents the largest improvement in angular resolutions of radio telescopes, far superior to those achievable with optical observations. The development and use of self-calibration techniques in the 1970s—a decade after the development of the VLBI, with a base line of thousands of kilometers—allowed the construction of images with a milliarcsecond resolution.

During the twentieth century the use of radio telescopes led to many interesting results and important discoveries not only in astrophysics, with the discovery of superluminal radio sources moving at speeds apparently faster than the speed of light and the observation of pulsars and quasi-stellar radio sources (commonly known as quasars); but also with the possibility of addressing cosmological questions, such as studying the cosmic microwave background to investigate the age and evolution of the universe. Current projects that radio astronomers are working on, are in the areas of millimeter and submillimeter interferometry and water vapor radiometry (for dealing with atmospheric phase fluctuations). These, combined with advanced digital processing techniques, are expected to yield results on the bases of which future goals in astronomy research could be formulated.

PEDRO RUIZ CASTELL

Further Reading

Hanbury-Brown, R. *Boffin: A Personal Story of the Early Days of Radar, Radio Astronomy and Quantum Optics*. Institute of Physics Publishing, Bristol, 1991.

Kellermann, K.I. and Moran, J.M. The development of high-resolution imaging in radio astronomy. *Ann. Rev. Astron. Astrophys.*, 39, 457–509, 2001.

Malphrus, B.K. *The History of Radio Astronomy and the National Radio Astronomy Observatory: Evolution Toward Big Science*, 1996.

Pawsey, J.L. and Bracewell, R.N. *Radio Astronomy*. Clarendon Press, Oxford, 1955.

Smith, R.C. *Observational Astrophysics*. Cambridge University Press, Cambridge, 1995.

Sullivan III, W.T. *The Early Years of Radio Astronomy*. Cambridge University Press, Cambridge, 1984.

Verschuur, G.L. *The Invisible Universe Revealed: the Story of Radio Astronomy*. Springer Verlag, New York, 1987.

Telescopes, Space

The ability to place astronomical telescopes and other detectors in space has given rise to the term "space astronomy," which describes astronomy performed from space rather than earthbound observatories. Its main advantage over terrestrial astronomy is its elimination of the deleterious effects of the earth's atmosphere, which not only absorbs some wavelengths of radiation but also distorts the images of stars and other objects of interest. Indeed, astronomy at some wavelengths (e.g., far infrared, submillimeter and x-ray astronomy) has advanced only through space astronomy, simply because the radiation is absorbed by the earth's atmosphere.

Although the term "space telescope" may refer to a module or payload on a space station, it is more commonly applied to a dedicated astronomical satellite or space observatory. There have been literally hundreds of space astronomy missions, observing in a wide variety of wavelengths from the radio end of the spectrum to x-rays and gamma rays.

The first space-based observatory was the National Aeronautics and Space Administration's (NASA's) Orbiting Solar Observatory, OSO-1, which was launched in March 1962 to expand on the work of the Explorer spacecraft series. OSO-1 operated for almost two years, during which time it transmitted data on more than 140 solar flares. The data collected from its 13 experiments was tape-recorded on board the satellite and transmitted to earth in a 5-minute period on each orbit (a communications method known as store-and-forward).

Previous science spacecraft had been designed as individual instrument carriers with no attempt at standardization, but the nine OSO spacecraft were based on a standardized observatory platform on which a variety of different payloads could be mounted—a design method that has become the norm for most types of spacecraft. Thus, although OSO-1 weighed only 206 kilograms at launch, by the time OSO-8 was launched in 1975, the standardized platform was supporting a spacecraft weighing 1052 kilograms.

The OSO platform was spin-stabilized by rotating the base section, using nitrogen thrusters

known as "gas jets," while the upper section, which contained the pointing-dependent part of the astronomical payload, remained pointing towards the sun. The spacecraft's spin axis was kept perpendicular to the solar vector by magneto-torquer coils in the base section, which aligned themselves, and thus the spacecraft, with the earth's magnetic field. Meanwhile, a gyroscope in the upper section acted as a memory to ensure that the spacecraft acquired the sun's light quickly on each orbit after emerging from the earth's shadow.

The early OSO launches were followed in 1968 by NASA's Orbiting Astronomical Observatory (OAO), which carried no less than eleven telescopes, enabling it to observe stars in the infrared, ultraviolet, x-ray and gamma-ray parts of the spectrum. Telescope mirrors up to 96 centimeter in diameter could be mounted inside the satellite's cylindrical core. Having realized that a spin-stabilized platform could not provide the stability required for accurate astronomical observations, a three-axis stabilized platform was designed for OAO. Its attitude was controlled by a number of rotating wheels mounted inside the spacecraft and aligned with each of its three axes, a type of stabilization system now used for the majority of manned and unmanned spacecraft.

The best-known space observatory is NASA's Hubble space telescope (HST), an optical telescope of the Cassegrain type named after the American astronomer Edwin Powell Hubble. It incorporates a number of camera and spectrometer payloads tuned to various frequency bands, and was designed to be launched to, serviced in, and eventually retrieved from, low-earth orbit by the Space Shuttle. Some indication of the progress made between the OSO series and the HST is provided by the increase in mass: the HST weighed some 11,250 kilograms at the time of its launch in April 1990.

The pointing accuracy of space telescopes has also increased markedly since OSO-1, which was accurate only to about 1 arc-minute. OSO-8, for example, equaled the arc-second accuracy of terrestrial telescopes and the three-axis stabilized OAO series reached 0.03 arcsecond. For comparison, the HST has a pointing stability of 0.007 arcseconds and the European Space Agency's (ESA's) Hipparcos astrometry satellite has an incredible 0.001 arcsecond stability.

Prior to launch, the capabilities of the HST's revolutionary payload were well publicized. Its 800-kilogram, 2.4-meter-diameter primary mirror, for example, would enable the telescope to resolve something the size of a small coin from a distance of 20 kilometers and detect the light of a firefly from about 16,000 kilometers. Unfortunately, it was discovered after launch that the mirror had been ground incorrectly as a result of a measurement error, to the extent that it was 2 micrometers (0.002 millimeters) too flat at the edges.

A servicing mission to correct the mirror's spherical aberration was conducted in December 1993, replacing one of HST's five main payloads with an optical instrument known as COSTAR (corrective optics space telescope axial replacement), which deployed five pairs of small corrective mirrors between the primary mirror and three of the remaining payloads. Subsequent servicing missions have changed or upgraded the original payloads and HST has made many successful observations and discoveries, including the derivation of a new value for the Hubble constant (the constant of proportionality between the recession speeds of galaxies and their distances from each other) and thus an improved estimate of the age of the universe. In fact, the HST is so popular among astronomers that observing time, which has to be booked in advance, is many times oversubscribed.

A selection of some of the more recent space-based telescopes is provided in Table 1. At the time of writing, several advanced missions are being planned by space agencies, including NASA's Next Generation Space Telescope, the European Space Agency's Integral gamma ray source mapper, and a number of infrared telescopes.

See also **Space Shuttle; Space Stations; Telescopes, Radio; Telescopes, Computer-Controlled Mirrors**

MARK WILLIAMSON

Further Reading

Cornell, J. and Gorenstein, P., Ed. *Astronomy from Space: Sputnik to Space Telescope*. MIT Press, Cambridge, MA, 1983.

Field, G. and Goldsmith, D. *The Space Telescope: Eyes Above the Atmosphere*. Contemporary Books, Chicago, 1989.

Goodwin, S. *Hubble's Universe: A New Picture of Space*. Constable, London, 1998.

Kopal, Z. *Telescopes in Space*. Faber & Faber, London, 1968.

Lawton, A.T. *A Window in the Sky: Astronomy from Beyond the Earth's Atmosphere*. Pergamon Press, New York, 1979.

Leverington, D. *New Cosmic Horizons: Space Astronomy from the V2 to the Hubble Space Telescope*. Cambridge University Press, Cambridge, 2000.

Peterson, C.C. and Brandt, J.C. *Hubble Vision: Astronomy with the Hubble Space Telescope*. Cambridge University Press, Cambridge, 1995.

Table 1 Selected Space Telescope Missions.

Spacecraft nation/agency	Launch date	Waveband(s)/mission
Hipparcos ESA	Aug 1989	Accurate measurement of star positions
COBE NASA	Nov 1989	Detect and map cosmic background radiation
HST NASA	Apr 1990	Visible, IR, UV imaging telescope
ROSAT Germany	Jun 1990	X-ray and extreme UV all-sky survey
GRO/Compton NASA	Apr 1991	Gamma ray observatory
ISO ESA	Nov 1995	Infrared observatory
SOHO ESA	Dec 1995	Solar and heliospheric observatory
SAX Italy	Apr 1996	X-ray observatory
Muses B/Haruka Japan	Feb 1997	Galactic radio source mapping
Chandra NASA	Jul 1999	X-ray observatory
XMM/Newton ESA	Dec 1999	X-ray spectroscopy & mapping

Television, Beginning Ideas (late Nineteenth and Early Twentieth Century)

In 1873 Willoughby Smith discovered the photoconductive property of selenium, the changing of electrical conductance with light falling on its surface. This property was utilized in many of the early schemes for television (the word dates from 1900) until the development of suitable amplifiers and photoemissive cells made selenium cells obsolete in the 1920s.

The photoconductive property of selenium was easily demonstrated and, in the following decade there was an expectation that "distant vision" would soon be a reality. This expectation was probably encouraged by the work in 1880 of Alexander Graham Bell and Charles Tainter on the photophone, a communication device that permitted sounds to be transmitted over a distance by means of a modulated beam of sunlight aimed at a selenium cell. The simplicity of the photophone and the lack of effort involved in its development, together with Bell's invention of the telephone in 1876, which enabled "hearing by electricity" to be readily implemented, stimulated workers in their quest to achieve "seeing by electricity." Several suggestions for "telectroscopes" were put forward in the 10-year period following Willoughby Smith's 1873 discovery.

The earliest ideas for these were based on notions that had been advanced from 1843 for picture telegraphy systems, but from around 1883 more appropriate ideas began to be proposed. In a basic system of television the light values of each elementary unit area of an illuminated scene or object are determined by an analyzing scanner-photoelectric cell arrangement and the amplified varying electrical signal is transmitted to the receiving apparatus. Here, the signal is again amplified and applied to an electrically controlled varying light source system so that, by means of a synthesizing scanner, an image of the original scene or object can be reproduced.

The implementation of such a system was not easy and more than 50 years would elapse, before John Logie Baird demonstrated a rudimentary form of television. Nevertheless, the work of the nineteenth century television pioneers was not wholly unproductive and, by the end of the century, some of the basic system components needed to implement a television scheme had been put forward. The ideas of Paul Nipkow (1883), Lazare Weiller (1889), and Marcel Brillouin (1894) led to television scanners—the apertured disk, the mirror drum, and the lens disk, respectively—which were widely utilized in the 1920s and 1930s. In addition Wilhelm Hallwachs' work on the photoelectric effect in 1888 followed by the detailed

investigations from 1889 of Julius Elster and Hans Geitel on photoelectricity, together with the fundamental researches which were being undertaken, contemporaneously, on the conduction of electricity in gases and in vacuums were important contributions that were to play a vital part in the progress of television. The latter work led to the invention, by Karl Ferdinand Braun in 1897, of the cathode ray tube (CRT) as a practical laboratory instrument.

Following Braun's publication, the development and use of cathode ray tubes was pursued by several investigators, and so it was perhaps inevitable that it would be incorporated into a television system. Boris Rosing, in Russia in 1908, was the first person to engage in work on television using a CRT receiver, although prior to this date Max Dieckmann and Gustav Glage, in 1906, had developed a "method for the transmission of written material and line drawings by means of cathode ray tubes," In his work, Rosing was assisted by his student Vladimir Zworykin who later, from 1923, and in the U.S., tried to evolve an all-electronic television system. He described his "kinescope" (CRT) display tube in 1929 and his "iconoscope" (CRT) camera tube in 1933.

On 9th May 1911 Rosing recorded in his notebook, "... A distinct image was seen [on the screen of the CRT] for the first time consisting of four luminous bands." In the same year Alan Archibald Campbell-Swinton elaborated on his earlier (1908) ideas, founded on a cathode ray camera tube and a cathode ray display tube, for an all-electronic television system. These ideas influenced James McGee who, from 1932, led the team at Electric and Musical Industries (EMI), which developed the emitron camera tube, the British equivalent of the iconoscope.

During the period 1912–1922 only a few new schemes for television were advanced. This situation changed greatly from 1923 because of technical advances in electronics. In 1904 J. Ambrose Fleming invented the diode valve. Two years later Lee de Forest invented the Audion (triode) valve and in 1912 discovered that the valve, in addition to having applications in detecting and amplifying circuits, could be utilized in an oscillator to generate electromagnetic waves.

World War I gave an impetus to the use of valves in signaling systems, and so stimulated developments in circuit and radio techniques, so that by 1918 triodes could be manufactured to cover wide power and frequency ranges and were suitable for both receiving and transmitting purposes. Consequently, by 1920 the time was opportune for the establishment of sound broadcasting: the radio systems were available and public demand was growing. The growth of commercial radio telephony, domestic broadcasting, and facsimile transmission influenced the progress of television. By the early 1920s, all the basic components of a rudimentary television system appeared to be at hand; and so, from around 1923 determined efforts to advance television were being made in the U.K., the U.S., France, Germany and elsewhere. At first these endeavors were mainly those of individuals—Baird in the U.K., Charles Jenkins in the U.S., Edwin Belin of France and Denes von Mihaly, a Hungarian working in Germany—working in isolation from others, but from 1925 this situation changed. Bell Telephone Laboratories in New York began an ambitious program that led to an impressive demonstration of the first U.S. television in April 1927, by wire from Washington D.C. to New York. Later, General Electric, Westinghouse Electric and Manufacturing Company, and the Radio Corporation of America (RCA), of the U.S.; Fernseh AG and Telefunken of Germany; the Marconi Wireless Telegraph Company, the Gramophone Company, Electric and Musical Industries (EMI), in addition to Baird's company in the U.K. and its subsidiaries, and others, all carried out experimental investigations in the television field. Baird made transatlantic transmissions of television signals by shortwave radio from London to New York in February 1928, and began experimental broadcasts in collaboration with the British Broadcasting Corporation (BBC) in September 1929.

In the U.K., the determined efforts of EMI led, on November 2, 1936, to the inauguration of the world's first, public, regular high-definition television service from studios and transmitters at Alexandra Palace, north London.

See also **Iconoscope; Television, Electromechanical Systems**

RUSSELL W. BURNS

Further Reading

Burns, R.W. *Television: An International History of the Formative Years*. Institute of Electrical Engineers, London, 1998.

Abramson, A. *The History of Television, 1880 to 1941*. McFarland, Jefferson, NC, 1987.

Burns, R.W. *British Television, the Formative Years*. Institute of Electrical Engineers, London, 1986.

Burns, R.W. *John Logie Baird, Television Pioneer*. Institute of Electrical Engineers, London, 2000.

Shiers, G. *Technical Development of Television*. Arno, New York, 1977.

Wilson, J.C. *Television Engineering*. Pitman, London, 1937.
Zworykin, V.K., and Morton, G.A. *Television: The Electronics of Image Transmission*. Wiley, New York, 1940.

Television, Cable and Satellite

In the early days of television experiments, transmission of signal was by wire connections and transmissions over phone lines. Radio communications came later. For example, the demonstration in April 1927 by Bell Telephone Labs between Washington, D.C. and New York City was transmitted by both wire and radio, and John Logie Baird also transmitted television between London and Glasgow using conventional telephone lines in 1927. All prewar British television outside broadcasts used post office telephone lines to get the video signal back to Alexandra Palace prior to radio broadcast. However, consumers received their signal only by radio, and television distribution by cable did not start until the 1950s, and even then in only a limited way.

During World War II, television broadcasts halted in Europe, but they carried on in the U.S., although in a reduced sense. After the war, television broadcasts in England restarted and by 1950 there were early attempts in London to use existing radio relay cables to distribute television. Radio relay was used in England, Germany, and Russia to distribute radio from a master antenna via cable to blocks of houses. By 1953, cable systems were being designed and installed elsewhere in England, still on the radio relay principle.

While there is dispute over who built the first U.S. system, the National Cable Telecommunications Association (NCTA) has given credit to Ed Parsons, who set up a local cable system in Astoria, Oregon. Seattle radio station KRSC had announced it would build television station KRSC-TV, now KING-TV. Parsons' wife had seen television in 1947 at a convention of the National Association of Broadcasters (NAB), and remarked that she would like to have television at home. When KRSC-TV went on the air November 25, 1948, Parsons had an antenna and booster installed on the roof of the eight-story John Jacob Astor Hotel, connected with twin lead to his penthouse apartment a short distance down Commercial Street. By New Year's Day 1949 he had run a feed to Cliff Poole's music store across the street from the hotel, using coaxial cable, and later to the whole town. Early cable transmission was by ordinary cable, but coaxial cable began to be used from 1941 in the U.S. and the 1950s in the U.K.

Cable TV developed as a medium for distributing or relaying broadcast signals to outlying communities which could not get good reception of broadcast signals due to being out of range of the transmitter. Early cable operators such as those in Pennsylvania in 1949 were mostly local entrepreneurs operating community antenna television or CATV, which consisted of a tall antenna with a repeater station and amplifier, connected by wire to a few homes. From about 1953, operators began to build microwave relays to bring in distant television signals. By 1961 there were 700 CATV systems in the U.S., though the industry soon coalesced into a few large operators called MSOs, or multiple system operators, as entrepreneurs began to consolidate small networks into larger operations. The industry continued to be primarily a master antenna service operating outside of major metropolitan areas until 1975, when Home Box Office (HBO) inaugurated satellite distribution of its pay movie service. The inaugural broadcast occurred on 1 October 1975 (Philippine time), with the transmission of the Ali–Frazier heavyweight prize fight from the Philippine Islands (the "Thrilla from Manila") to cable television systems in Florida and Mississippi. HBO then began to sell its satellite-delivered movie service to cable operators, who in turn sold it to subscribers. Shortly after, a number of cable networks, both pay and advertiser-supported, began to appear. With these new sources of programming not available off the air, cable television began to penetrate larger cities.

Communications satellites receive television signals from a ground station, amplify them, and relay them back to earth. Satellite distribution of television began in 1962. On 11 July 1962 satellite

dishes at the Radome in Pleumeur-Bodou, France and British Telecom's Goonhilly Earth Station, in Cornwall, U.K. received the first transatlantic transmission of a television signal from a twin station in Andover, Maine, in the U.S. via the TELSTAR satellite. TELSTAR had a low, elliptic orbit and was only usable for three or four 40-minute periods in each 24 hours. It delivered satellite television during its seven months in orbit.

The satellites used today for cable and broadcast program distribution are in geosynchronous orbit, so that they appear to be stationary in space, affording the use of relatively low-cost fixed receiving antennas. The early antennas were 10 meters in diameter, but antennas smaller than 3 meters can be used today. The original downlink (satellite-to-ground) was the so-called C band, 3.7 to 4.2 gigahertz. Analog frequency modulation was used, and remains in declining use today. Some newer satellites use downlinks in the Ku band, about 11.7 to 12.2 gighertz. Most newer satellite links use digital transmission of MPEG-2 (Motion Picture Experts Group) compressed video. This permits more programs to be transmitted in the same bandwidth, and permits use of smaller earth station antennas.

Cable television systems start at ''headends,'' where signals are brought together from a number of sources. Headends supply signals directly to subscribers located close by, and supply signals to more distant subscribers using fiber-optic cable. Figure 3 illustrates a modern cable television system distributing signals to a large metropolitan area. The headend supplies signals to a number of ''hubs.'' Radiating from each hub are a number of fiber-optic cables, each connecting to one or more ''nodes.'' The nodes convert signals from optical to electronic form, where they are distributed via coaxial cable (coax) and radio-frequency amplifiers, to individual homes. A portion of the signal is removed to send to a group of homes, at a ''tap.'' The tap is a passive (nonpowered) device that draws a predefined portion of the signal from the cable and sends it to one or more homes. Homes are attached to the tap by way of ''drop'' cable, smaller, flexible coaxial cables.

Signal strength is reduced as the signals travel through the coaxial cable. The signal strength reduction is due to two phenomena: (1) every time some of the signal is taken out of the cable at a tap, the remaining signal going downstream is weakened by the signal removed to supply the

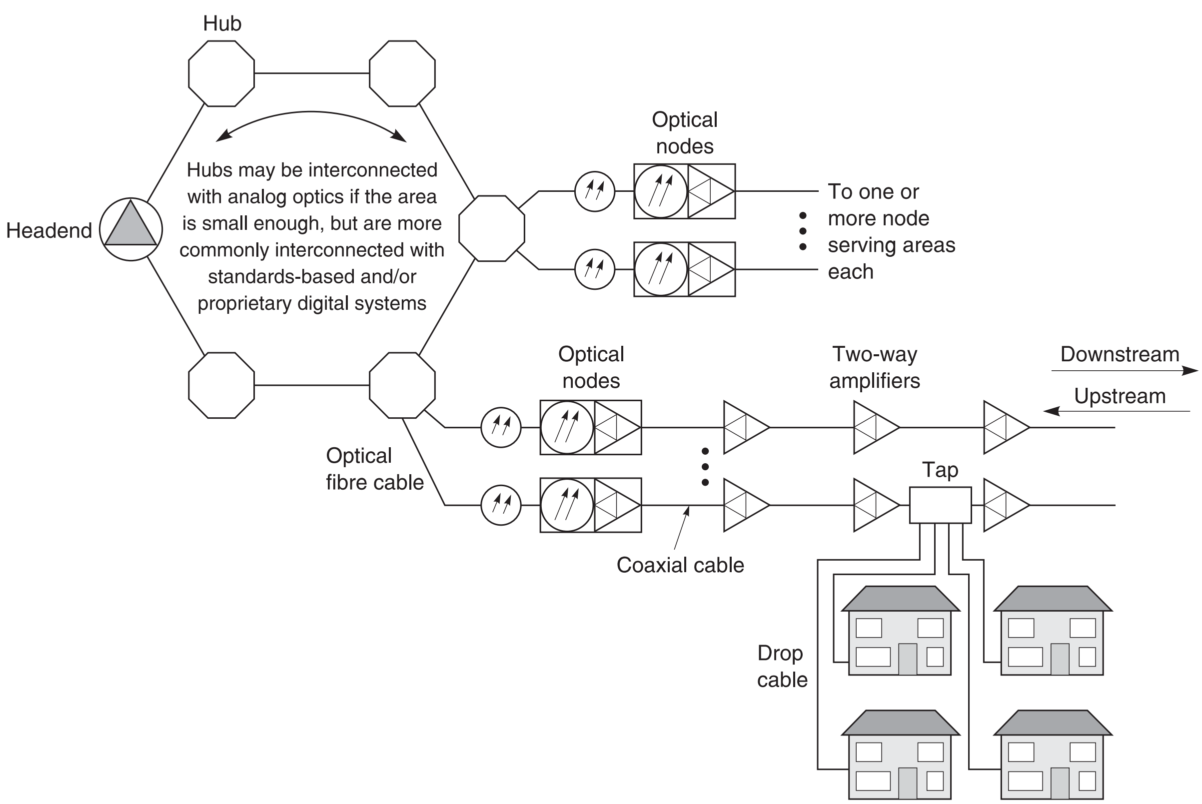


Figure 3. Traditional cable television program distribution.

homes attached to that tap; and (2) the coaxial cable itself causes signal loss due to resistance in the conductors and losses in the dielectric insulator between the conductors. Because of these losses, amplifiers must be used periodically to increase signal strength.

Traditional cable television program distribution is downstream only (see Figure 3). However, many modern services require two-way communications. These services include cable modem and telephone service, as well as interactive video. Bidirectional amplifiers facilitate this two-way communications by amplifying signals flowing in both directions. In order to separate the upstream and downstream signals, it is necessary to transmit them in different frequency bands. North American practice is to use the band from 54 megahertz to as high as 870 megahertz for downstream transmissions, and the frequency band from 5 to 42 megahertz for upstream transmission. Other parts of the world use slightly different frequency plans. Each amplifier uses "diplex filters" to separate these bands at both the input and output of the amplifier, then uses separate amplification circuits for the two frequency bands. Separate fiber strands usually carry different direction signals between the hub and the optical node.

Each amplifier adds distortion and noise to the signal, so the number of amplifiers used in any one signal path (called a cascade) must be limited. Also, since the amplifiers tend to be somewhat trouble-prone due to their complexity, reliability suffers as the cascade is lengthened. In early systems, amplifier cascades of 20 were common, and cascades of 50 amplifiers were sometimes used. Today, with the introduction of fiber optics, the number of amplifiers in a cascade has been reduced to typically six, with a strong trend toward fewer amplifiers.

Fiber-optic cables are being brought deeper into the cable plant to facilitate this reduction in the number of amplifiers in cascade. Long distances can be achieved with fiber-optic transmission (Figure 4). This figure plots the loss per kilometer for two sizes of coax, and also for fiber-optic cable. The loss on the vertical axis is measured in decibels (dB), with lower numbers (less loss) being better. This lower loss of fiber-optic cable permits long distances between the hub and the node, with no amplification being needed in most cases. Optical amplification is practical and is used where needed.

The systems described so far are called hybrid fiber-coax (HFC) systems. The latest trend is toward either fiber-to-the-curb (FTTC) or fiber-to-the-home (FTTH) systems, neither of which use

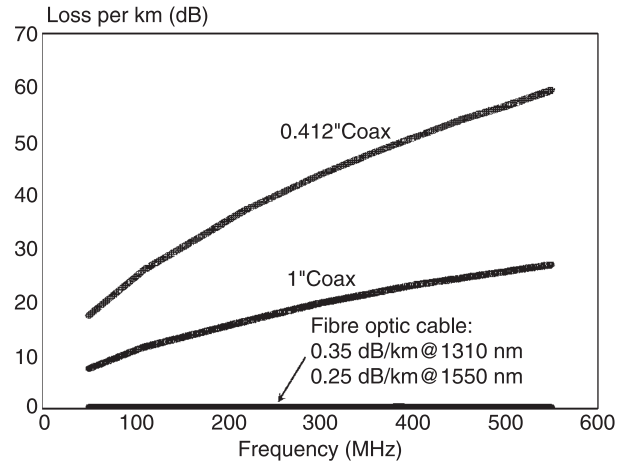


Figure 4. Signal loss per kilometer measured in decibels in fiber-optic cable and two sizes of coaxial cable. Lower numbers indicate less loss and therefore permit long distances between hub and node with no amplification needed in most cases.

RF amplifiers. In FTTC systems, the fiber is brought to a location that serves between four and sixteen homes, and broadcast services are distributed over coax from there. Telephone service is distributed from the end of the fiber to homes, using conventional telephone cable. Data is distributed on data cables, or in some cases may use the same cable as the phone service uses.

FTTH systems bring a fiber cable to the side of the home, where it terminates in a conversion device that supplies video, voice, and data signals for use in that one home. FTTH systems can be divided into passive systems, which have no active (amplification and signal processing) devices at all between the hub and the home (PONs, or passive optical networks); and active systems, which have one active device. The active systems can achieve much longer reach between the hub and the home, and can use lower cost optical components. These systems are beginning to be deployed at the time of writing, in competition with conventional HFC plants. These new architectures offer superior reliability, higher quality signals, and much higher data rates than are practical with HFC networks.

Signal distribution employs "frequency division multiplexing," or FDM, in which each signal is assigned a frequency band at which it is transmitted. The frequency band is called a channel. The signal is modulated, or impressed on, a carrier frequency in that assigned channel. For analog transmissions, one program is assigned to each channel, whereas for digital transmissions, several

programs may be transmitted simultaneously on one channel. Simultaneous transmission of multiple programs is done using "time division multiplexing," or TDM. In TDM, a portion of one program is transmitted, followed by a portion of another and so on. The transmission is fast enough that all programs appear to the user to be transmitted simultaneously. At the television, when the viewer selects a channel, he or she is tuning the television to that carrier frequency, and the television is recovering, or demodulating, the channel.

In HFC networks, data and telephone traffic are also transmitted using a similar combination of FDM and TDM. In FTTC and FTTH systems, television programs may be transmitted using FDM as on HFC networks, but data and telephone traffic normally use base-band TDM, in which they are not modulated onto carriers, but rather are time division multiplexed and used to turn optical transmitters on and off at a fast rate. Video programs may also be transmitted this way, although it remains more common to use FDM to transmit video programs.

See also **Optoelectronics; Satellites, Communication**

JAMES O. FARMER

Further Reading

Adams, M. *Open Cable Architecture*. Cisco Press, Indianapolis, 2000.

Burns, R.W. *Television: An International History of the Formative Years*. Institute of Electrical Engineers, London, 1998.

Ciciora, W., Farmer, J. and Large, D. *Modern Cable Television Technology: Video, Voice and Data Communications*. Morgan Kaufmann, San Francisco, 1999.

Easton, K.J. *Thirty Years in Cable TV: Reminiscences of a Pioneer*, Pioneer Publications, Mississauga, Ontario.

Farmer, J. *Fiber-to-the-Home: Why it's Inevitable*. PBIMedia, Potomac, MD, 2001.

Hecht, J. *City of Light: History of Fiber Optics*. Oxford University Press, New York, 1999.

Junkus, J.J. and Sawyer, M.J. *Digipoints*, vols. 1 and 2. Society of Cable Telecommunications Engineers, Exton, 1998 and 2000.

Laubach, M., Dukes, S. and Farber, D. *Breaking the Access Barrier: Delivering Internet Connections over Cable*. Wiley, New York, 2001.

Raskin, D. and Stoneback, D. *Return Systems for Hybrid Fiber/Coax Cable TV Networks*. Prentice Hall,. Upper Saddle River, NJ, 1997.

Taylor, A.S. *History Between Their Ears*. Cable Center, Denver, 2000.

Walker, P., Ed. *Milestones: A History Of Cable Television*. National Cable Television Center And Museum, Denver, 1998.

Useful Websites

The Cable Center, Cable History: http://www.cablecenter.org/history/index.cfm

Television, Color, Electromechanical

Prior to 1861 photographs in color could only be reproduced by hand painting daguerreotypes, which were direct-image photographs produced on a silver-coated copper plate. In 1855 James Clerk Maxwell suggested a method for creating a color image. If a scene or object were photographed through red, green, and blue filters separately to obtain three negatives and the three positives prepared from them were projected with the images in alignment onto a screen by means of three lanterns, fitted with the appropriate red, green, and blue filter in front of the projection lens, a colored image would result. Maxwell demonstrated the correctness of his notion at a meeting of the Royal Institution in London in May 1861.

John Logie Baird adapted Maxwell's concept to display colored television images on 3 July 1928, for the first time anywhere. His sending and receiving apparatus utilized Nipkow disk scanners, each having three spiral sets of apertures. The three sets were covered separately by red, green, and blue filters. At the sending end of the television link the scene or object was analyzed, sequentially, into its three primary color components, and at the receiving end the video signals corresponding to these were used to reconstitute three primary color images in register. Because of the persistence of vision, the final image was colored.

On 27 June 1929 Bell Telephone Laboratories (BTL) gave a demonstration in New York of color television using Nipkow disk scanners and transmitting stills. However, whereas Baird used a single scanner having three spirals, at each end of his television link and a single transmission channel, BTL employed a single scanner, having a single spiral of apertures, at each end of the television link, and three transmission channels. The red, green, and blue color contents of the scene or object being televised were transmitted simultaneously and not sequentially. An advantage of this method was that the same scanning disks and motors, synchronizing equipment, and circuits were applicable as in the monochrome television scheme. Neither scheme led to color television broadcasting services.

On 4 February 1938 Baird gave a public demonstration of television in the Dominion Theatre, London, at which high-definition images

of about 3 by 2.7 meters, were shown in color, the television signals being received by radio, using a wavelength of 8.3 meters, from the Crystal Palace transmitter, about 16 kilometers away

The transmitting apparatus consisted of a 20.3-centimeter-diameter mirror drum, provided with 20 mirrors rotating at 6000 revolutions per minute (rpm). These mirrors reflected the scene to be transmitted, through a lens, onto a 500-rpm Nipkow type disk provided with 12 concentric slots positioned at different distances from the disk's axis. Each of the slots was covered by a color filter, blue–green and red being used alternately. By these means the fields given by the 20-mirror drum were interlaced six times to give 120-line picture signals repeated twice for each revolution of the disk.

At the receiver a similar system of rotating mirror drum and disk was employed together with a high-intensity arc lamp source. The light intensity was modulated; that is, made to fluctuate, in exact conformation to the variations from the transmitting end, as it passed through a modified Kerr cell connected to the television receiver.

Another demonstration was given by Baird, in his private London laboratory, on 27 July 1939. However, Baird utilized a cathode ray tube (CRT) at the receiver in place of the mirror drum and slotted disk. A rotating disk having 12 circular filters, alternately red and blue–green, was positioned in front of the screen of the CRT.

The following year Baird, on 17 September, patented a method of color television that enabled him to demonstrate, in April 1941, 600-line color television. Again the two-color principle was adopted but the sending-end scanner was now a CRT of the type that had been employed as a projection unit for cinema television. Both the sending-end and receiving-end CRTs had two-color filter disks rotating in front of their screens.

Further development of this system allowed Baird to demonstrate colored stereoscopic images to the press on 18 December 1941 (Figure 5). At the transmitter, the primary scanning beam of light, after having passed through one of the color filters and the projection lens, was divided by a system of two pairs of parallel mirrors into two secondary beams spaced apart by a distance equal to the average separation of an observer's eyes. By means of the revolving shutter disk, the scene was scanned alternately by each secondary beam.

The receiver included a color disk identical to that of the transmitter and a revolving shutter, both synchronized to the corresponding disks at the sending end of the system. Hence each eye of the viewer alternately observed red, green, and blue images. The shutters used differed at the transmitting and receiving ends to minimize flicker.

Baird's only competitor in the color television field by 1940–1941 was the Columbia Broadcasting System (CBS). On September 4, 1940 in New York, CBS engineer Peter Goldmark demonstrated equipment that comprised a Farnsworth image dissector tube camera and associated three-color filter disk rotating at 1200 rpm at the sending end; and a 23-centimeter CRT and another identical rotating filter disk at the receiving end. A 343-line image 14 by 18.5 centimeters was displayed.

From early 1942 Baird began to consider nonmechanical methods of color television and patented one version on 13 May 1942. He adapted the Thomas system of color cinematography in which an optical unit in the camera automatically produces images on a black and white film of the red, green, and blue components of a scene. At the receiver an identical optical unit combines the three cine film images to display a colored image. In the television version of this system the film camera and projector were replaced by an electronic camera and a projection CRT.

Two years later, on 16 August 1944, Baird gave a demonstration to the press of his telechrome tube—the world's first multigun, color television display tube. The tube employed either double or triple, separate and independent, electron guns and multiple fluorescent screens depending upon whether two- or three-color reproduction was required. Only the two-gun version of the telechrome tube was demonstrated.

In 1944 it was clear that electromechanical methods of color television would not be viable in a domestic situation—although CBS persisted with its scheme—and by 1950 all electronic frame-, line-, and dot-sequential color television systems were being considered.

See also **Television, Electromechanical Systems; Television: Color, Electronic**

RUSSELL W. BURNS

Further Reading

Abramson, A. *The History of Television, 1880 to 1941*. McFarland, London, 1987.

Burns, R.W. *Television: An International History of the Formative Years*. Institution of Electrical Engineers, London, 1998.

Burns, R.W. *John Logie Baird, Television Pioneer*. Institution of Electrical Engineers, London, 2001.

Ives, H.E. Television in color. *Bell Laboratory Record*, 7, 439–444, July 1929.

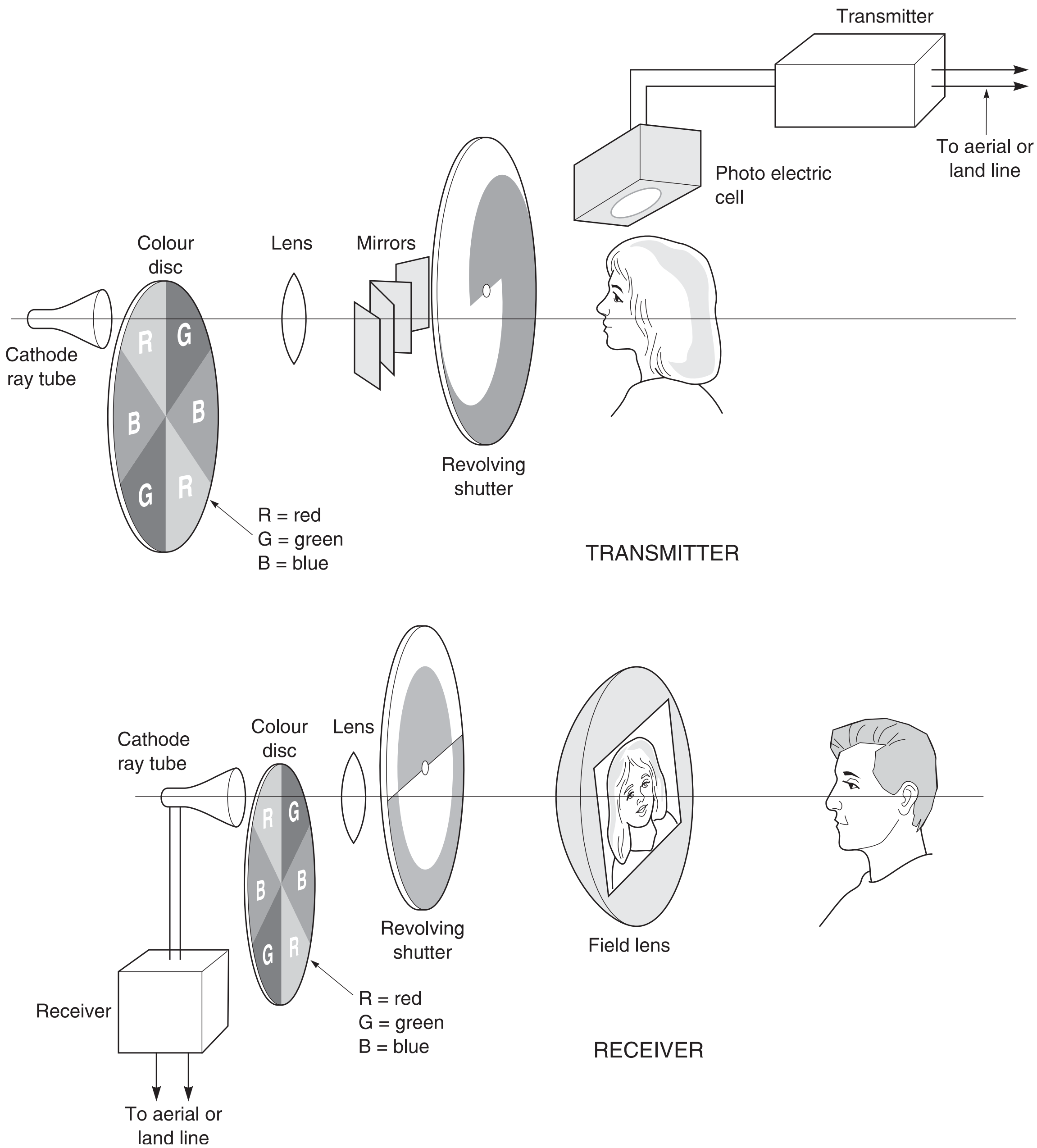


Figure 5. Schematic diagrams showing the layout of the apparatus that Baird used to show stereoscopic color television.
[*Source: Electrician, December 1941, p. 359.*]

McLean, D. F. *Restoring Baird's Image*. Institution of Electrical Engineers, London, 2000.

Television: Color, Electronic

At the end of the twentieth century, some 900 million people watched electronic color television around the world. Americans owned almost as many color television receivers as indoor toilets. In a generation the medium supplanted wireless broadcasting and cinema as the world's premier source of news and entertainment.

Although John Logie Baird had worked on an all-electronic color television from 1942, the development of this technology after World War II can be assigned to one organization. Under the sponsorship of chief executive David Sarnoff, the Radio Corporation of America (RCA)

Laboratories undertook the research and development necessary to make color television a reality between 1945 and 1953. The staff also made the transmission standards and hardware compatible with those in place for monochrome television. After the Federal Communications Commission (FCC) approved standards based on this effort, RCA underwrote production, programming, and marketing efforts until American consumers began returning RCA's $100 million investment in the early 1960s. Other countries adopted this standard or adapted it with various modifications.

Until RCA's researchers took up the challenge after World War II, it appeared that the FCC would approve the Columbia Broadcasting System's (CBS) electromechanical field-sequential system of color television, which was incompatible with the monochrome system already in place. Supported by other RCA engineers, the staff at RCA Labs proposed the conceptual framework, established most of the principles and techniques, and demonstrated the technologies necessary for the electronic pickup, broadcast, and reception of color television in the home. This system was the foundation for analog broadcast standards internationally.

The laboratories' researchers had experimented with field-sequential color systems during the early 1940s. This research demonstrated the inherent limitations of the technology in terms of picture brightness, monochrome compatibility, and image scalability. RCA then committed to developing an all-electronic system. In 1947 the Labs demonstrated a color system using Alda Bedford's principle of "mixed highs." Because the human eye distinguishes changes in brightness but not color at high levels of detail, high-frequency components of the primary additive colors—red, green, and blue—could be blended into one signal. This reduced the amount of bandwidth used.

By the end of 1948 technical and regulatory pressures forced RCA's researchers to compress the bandwidth to the 6 megahertz used for monochrome broadcasting. In January 1949, the Labs' leadership committed to meeting that limit while maintaining equivalent resolution, brightness, and flicker. Monochrome signals would be displayed on color receivers, while color signals would be received without adjustment on monochrome receivers.

That September, the FCC began hearings on proposed formats for color television. For the next eight months, under the pressure of attention by the government, the media, and the competition, RCA Labs staff turned their concepts into demonstrations of principle. Mixed highs and Clarence Hansell's application of time-division multiplexing to color transmission permitted the compression of the televised signal into two components. One carried the luminance, or brightness, and the other the chrominance, or color information. A monochrome set would simply ignore the color component. John Evans and Randall Ballard's use of dot interlacing to scan the colors was combined with Bedford's color-sampling reference burst to synchronize frequencies between transmitter and receiver. Harry Kihn's "Kolor Killer" circuit enabled color receivers to show monochrome signals.

As for hardware, Richard Webb built color cameras using three image orthicons, the standard monochrome image tube developed at the labs. Each tube scanned an image in a primary color. Dichroic mirrors (which transmit certain wavelengths or colors of light while reflecting others) and electronic circuitry then blended the three primaries and sent the image to the amplifier. RCA began producing a commercial camera based on this design in 1952.

The weak point, technically and socially, was the receiver. RCA's "triniscope" used three cathode-ray tubes (CRTs) and dichroic mirrors to combine the primary colors and project the image on a screen. The sheer bulk of this design required that RCA develop a practical household alternative.

Harold Law's refinement of Al Schroeder's shadow-mask tube offered the best solution. Like a monochrome CRT, the screen's glow was based on the intensity of the electron beam and signal. But the color tube contained three electron guns, one for each primary color. A perforated mask next to the screen enabled the beams to strike the appropriate red, green, or blue phosphors clustered in thousands of triads. The intensity of the beam and the relative brightness of the phosphors determined the colors seen by the viewer. More than any other component of the system, the shadow-mask CRT made color television (and computers) a household technology.

In July 1950, the National Bureau of Standards (NBS) reported that RCA's system was demonstrably superior to CBS's system and offered the most "opportunity for improvement." Nonetheless the FCC rejected RCA's system in September 1950. The rest of the television industry had already started refining RCA's format and technologies. The second National Television System Committee (NTSC) was organized in January 1950. Members of the industry and the FCC agreed to develop with RCA a standard for electronic color television broadcasting and reception. The NTSC adopted RCA's system with one significant exception.

Hazeltine Corporation, an independent laboratory with a cross-licensing agreement with RCA, developed a different approach to analyzing the color signal. While RCA's staff understood this process as one of sampling the combination of brightness and color, Arthur Loughren and Bernard Loughlin represented the dot-sequential signal as a color subcarrier added to a monochrome-carrier wave. This change in perspective enabled engineers to treat color television with the traditional tools of monochrome video and sine-wave mathematics. In April 1950 Hazeltine demonstrated its concepts of constant luminance and "shunted monochrome" to RCA Labs, which adopted them to resolve outstanding display problems.

RCA unveiled its version incorporating this approach in June 1951; field testing began in February 1952. During this time Schroeder and Ray Kell reduced color fringing by adapting quadrature-amplitude modulation. Color was now transmitted on a wideband orange–cyan I axis and narrowband green–purple Q axis.

The FCC approved the NTSC standard on December 17, 1953. Japan, Canada, Central America, and some South American countries followed suit. In the 1960s, all of Western Europe except France adopted the German company Telefunken's NTSC variant, phase alternating lines (PAL); France persuaded Russia and Eastern Europe to use sequentiel colour avec mémoire (SECAM). PAL, adapted from NTSC, has phase reversal that avoids color errors resulting from amplitude and phase distortion of the color modulation sidebands during transmission. For many European television engineers, NTSC's receivers, which have manual tint and intensity controls, have such large consumer color control that NTSC is said to stand for "never the same color twice," while American engineers dub SECAM "system essentially contrary to the American method." Today, the North American NTSC system is still incompatible with France and Eastern Europe's SECAM and Western Europe's PAL, with different receivers required for each.

Until the mid-1960s color cameras continued to use image orthicons, when Philips's plumbicon tube supplanted them; RCA introduced solid-state charge-coupled devices (CCDs) to cameras in 1984. Color receivers continue to feature shadow-mask variants with Sony's 1968 Trinitron tube being the notable improvement.

See also **Television: Color, Electromechanical**

Alexander B. Magoun

Further Reading

Barnouw, E. *Tube of Plenty: The Evolution of American Television*, 2nd revised edn. Oxford University Press, New York, 1990. Best history of programming and its relationship with technology and regulation.

Brown, G.H. *And Part of Which I Was: Recollections of a Research Engineer*. Angus Cupar, Princeton, 1982. Thoughtful and humorous account of color standards competition by head of RCA's color program.

Fisher, D.E. and Fisher, M.J. *Tube: The Invention of Television*. Harcourt, New York, 1997; Counterpoint, Washington D.C., 1996. Relies on secondary sources for history of color.

Inglis, A.F. *Behind the Tube: A History of Broadcasting Technology and Business*. Focus Books, Boston, 1990. Best analysis of competing color systems and corporate motives for their support.

Schreiber, W.F. and Buckley, R.R. Introduction to color television—part I. *IEEE Proc.*, 87, 1, 173–179, January 1999. Argues that field-sequential color system could have worked with color tube and monochrome compatibility was a mistake.

Slotten, H. *Radio and Television Regulation: Broadcast Technology in the United States, 1920–1960*. Johns Hopkins University Press, Baltimore, 2000. Study of standards debates, including that over color, as struggle between those who understood and designed the systems and the political bodies that had to regulate it.

Useful Websites

IEEE History Center RCA Oral History Collection: http://www.ieee.org/organizations/history_center/oral_histories/oh_rca_menu.html

Interviews with RCA's Harold Law, Humbolt Leverenz, and Paul Weimer, who had basic roles in developing RCA's system.

Television, Digital and High-Definition Systems

The term "high definition" was first applied to television in the 1930s to compare all-electronic analog television (TV) systems with the older and partially mechanical systems then still used experimentally. Regular television service as inaugurated by the British Broadcasting Corporation (BBC) in 1936 (using 405 scanning lines) and in the U.S. in 1941 (525 lines) was often termed "high definition," as was a Dumont experimental system of 1939–1940 that briefly achieved 800 lines. Decades later, however, the term came to mean something quite different.

Originally restricted to analog technology, high-definition (usually defined as greater than 1,000 scanning lines, roughly equivalent to 35-millimeter film quality) television development focused on digital methods after 1990. By the early twenty-first century, however, a multiplicity of digital television

technical standards appeared more likely than world agreement on any single approach. Digital television was thus following in the footsteps of analog, where in the 1960s three different color TV systems developed more for economic and political than technical reasons: NTSC, National Television System Committee, from the U.S.; PAL, phase alternative (or alternating) line, from Germany; and SECAM, sequential colour avec mémoire, from France.

HDTV Origins

Modern era high-definition television research began in Japan in the late 1960s as Nippon Hoso Kyokai (Japan Broadcasting Corp.) engineers sought to improve on the NTSC 525-line standard. This effort expanded into full-time research into both video and audio aspects of an improved system about 1970. The resulting 1125 scanning-line analog "Hi-Vision" system was first demonstrated to American policymakers in early 1981, four years after American engineers had also begun to investigate HDTV's potential. Interlaced scanning and field or frame standards matched existing NTSC practice, but the picture's aspect ratio was widened to 16:9 rather than NTSC's 4:3 (to better approximate telecasting of widescreen motion pictures), and multiple sound channels were digital. Because the highly complex HDTV signal would require the equivalent of six (later four) NTSC 6-megahertz channels to transmit, however, Hi-Vision was designated for production and not transmission.

Concern over the likely high cost of any high-definition system led to development of several analog alternatives—dubbed "advanced" or "enhanced"—which, providing far fewer than 1,000 scanning lines, would require less spectrum space while still offering improved picture quality. Major U.S. manufacturers and trade associations established an Advanced Television Systems Committee (ATSC) in 1983 to coordinate American research efforts and conduct comparative tests. Some proponents argued for a gradual transition to HDTV through these intermediate systems. Research continued into developing HDTV systems that could be "downward compatible" (receivable in lower-quality analog form) on existing NTSC sets to ease the introduction of HDTV.

Several European countries had by the mid-1980s begun cooperative development of analog HDTV technology based on their 50 hertz electrical standard, seeking to avoid acceptance of a Japanese system, imports of which might wipe out their domestic consumer electronics industry as had happened earlier in the U.S. Europe soon focused on perfecting the MAC (multiplexed analog components) family of transmission standards. The Japanese concentrated increasingly on their MUSE (multiple sub-Nyquist sampling encoding) transmission standard which applied signal compression to force HDTV into only 8.1 (later 6) megahertz of bandwidth, thus making a broadcast service more likely. Both European and Japanese efforts focused on satellite-delivered HDTV, bypassing bandwidth-limited broadcast stations. Faced with substantial industry pressure not to pursue a satellite option, however, the U.S. focused more on a terrestrial broadcast service when the U.S. Federal Communications Commission issued its first inquiry concerning advanced modes of television in mid-1987. The FCC also soon made clear its preference for a true HDTV system, bypassing intermediate "advanced" or "enhanced" stages. What then appeared imminent, however, took more than a decade to even begin to achieve.

Digital Breakthrough

In mid-1990, General Instrument transformed the HDTV picture by announcing computer models of a proposed fully digital system ("DigiCipher") of high-definition television. Practical models were soon being tested in laboratories. Under industry and FCC pressure, several competing digital HDTV proponents merged the best parts of their systems into a so-called "Grand Alliance" in 1993 as laboratory and field testing continued. The FCC allotted an additional channel to each on-air station to encourage development of HDTV parallel to continuing analog transmissions. In late 1996 the commission adopted the Grand Alliance system as a formal set of standards. No less than 13 different video scanning modes were included, ranging from 480 to 1080 lines, and allowing either interlaced or progressive scanning. Agreement was reached on use of the MPEG-2 (Motion Picture Experts Group-2) set of compression tools to condense the HDTV signal by a ratio of 55 to 1, allowing it to fit into the 6-megahertz (8-megahertz in Europe) channels presently used for analog service.

In Europe, cooperation on what had become known as the D2-MAC standard collapsed early in 1992 for lack of sufficient demand and the expense of introducing the system. Growing European satellite television success was based on existing

analog formats, further undermining the attempt at a continent-wide digital standard. Faced with the digital transition elsewhere, Japan reduced its backing of analog MUSE, despite the fact the system was in regular NHK operation (though to few receivers given their very high cost) and commenced active work on a digital system.

Japan's introduction in 1991 of regular analog HDTV service (transmitted eight hours daily from a domestic satellite) was the world's first. Scheduled HDTV first aired in the U.S. when WRAL-TV (Raleigh, NC), began a limited but regular digital HDTV service five years later. Other stations slowly followed, most of them transmitting only a few hours a week. Costs of converting a station's facilities to full digital HDTV operation ranged up to $10 million or more. Stations in the ten largest markets began offering a few HDTV hours weekly in November 1988 and the system's use slowly expanded to smaller markets. By FCC rules, all stations were to be providing at least some HDTV service by May 2002.

Complicating the digital TV picture, American broadcasters by the turn of the century had become increasingly interested in an application of digital television that promised more immediate revenue—the ability to transmit four or more "standard" (or slightly degraded) NTSC channels rather than a single high-definition signal. They argued that this would avoid the huge expense of HDTV conversion (which showed little promise of creating additional industry revenue) while allowing broadcasters to better compete in a multichannel environment. In Europe and Japan, "enhanced" definition TV services were available by the late 1990s. Still, by 2002 it was increasingly clear that widespread use of HDTV would take far longer than proponents had earlier projected—easily another decade or more would pass before existing analog systems could be turned off.

Part of the delay and confusion is because, contrary to widespread opinion even within the industry, there is no single digital television standard for either production (there are actually thirteen) or transmission (where there are five, two of them for high-definition service). Another problem is confusion over nomenclature—when does high definition really mean *high* definition. Different players are pursuing different HDTV strategies using different standards, helping to confuse the marketplace. Finally, the cost of receivers (in the aggregate, far more costly than the industry's conversion costs) are off-putting to many would-be set buyers. Prices still averaged $3000 early in 2002 (though down from ten times that level a few years earlier). This last point means that "downconversion," or the ability of existing analog receivers to receive (with a conversion box) at least a semblance of "lower" definition television, is a hugely important aspect of what is now clearly going to be a slow transition.

Still, the promise of digital high definition remains strong. While the analog systems of today provide a video image made up of about 210,000 pixels (individual picture elements), digital HDTV images provide ten times that number. That image can be formatted in either a standard (three units high by four units wide) or wide-screen (nine high by sixteen wide) frame. The latter is closer to theatrical film's format, and thus avoids having to use either letterbox (leaving gray bars above and below the film) or pan-and-screen (editing the picture to better "fit" the television screen) techniques when showing films on television. A major controversy in developing digital standards has been whether to continue the use of interlaced picture scanning (standard in analog systems), or adopt the progressive scanning used on computer screens. The latter is better in many ways (smoother transitions and generally more capable), but requires more bandwidth. U.S. policymakers have allowed the use of either mode, leaving the eventual decision to individual broadcasters. This may force the manufacture of television receivers able to do both, at greater expense than a clear FCC standards decision one way or the other. Finally, the approved audio standard for digital television—the five-channel Dolby Digital (AC-3) surround sound system—compresses 5.1 channels of audio to a 640 kilobytes per second stream. While sophisticated, most current television production and transmission equipment cannot adequately handle it and some time will pass before consumers can receive the full advantage of this system.

CHRISTOPHER H. STERLING

Further Reading

Abramson, A. *The History of Television, Volume 2: 1941–2000*. McFarland, Jefferson, NC, 2002.

Benson, K.B., and Fink, D.G. *HDTV: Advanced Television for the 1990s*. McGraw-Hill, New York, 1990.

Brinkley, J. *Defining Vision: The Battle for the Future of Television*. Harcourt Brace, New York, 1997.

de Bruin, R. and Smits, J. *Digital Video Broadcasting: Technology, Standards, and Regulations*. Artech House, Norwood, MA, 1999.

Dupagne, M., and Seel, P.B. *HDTV: High-Definition Television-A Global Perspective*. Iowa State University Press, Ames, 1998.

Evans, B. *Understanding Digital TV: The Road to HDTV*. IEEE Press, New York, 1995.

Forrester, C. *The Business of Digital Television*. Focal Press, Boston, 2000.
Hartwig, R.L. *Basic TV Technology: Digital and Analog*. Focal Press, Boston, 2000.
Sudalnik, J.E. and Kuhl, V.A. *High-Definition Television: An Annotated Multidisciplinary Bibliography, 1981–1992*. Greenwood, Westport, CT, 1994.

Useful Websites

Advanced Television Systems Committee (ATSC) website: http://www.howstuffworks.com/framed.htm?parent=hdtv.htm&url=http://www.atsc.org/

Television, Electromechanical Systems

Television is a form of telecommunication for the transmission of signals representing scenes—the images of the scenes being reproduced on a screen as they are received or recorded for subsequent use. In monochrome television the luminance, but not the color, of an object is reproduced; in color television the reproduced picture simulates both the color and the luminance of the object.

As with a cine film, television consists of a series of successive images that are manifested by the brain as a continuous picture, because of the persistence of vision. Unlike a cine film, where the light values of each picture element of a scene are simultaneously recorded on a film, the light values of a televised scene are scanned in portions that are transmitted sequentially to the recording or display device, since in practice, only a single transmission link is used for a given television camera-display system. This restriction necessitates the utilization of a scanning device and a photosensitive cell at the transmitter to sample a two-dimensional image and convert it into an electrical signal, and another scanner and a display device at the receiver to synthesize or reconstitute the reproduced image from the transmitted electrical signals. Usually, the scanning process follows a left-to-right, and a top-to-bottom sequence as in the reading of a printed page.

Electromechanical television scanners date from c.1880. Although very many suggestions were advanced during the period 1880 to 1930, only the apertured disk scanner of Paul Nipkow (1884), the mirror drum scanner of Lazare Weiller (1889), and the lensed disk scanner of Marcel Brillouin (1891) were subsequently extensively used by experimenters. Of these types, Nipkow's disk was the simplest and the most versatile.

Figure 6 shows a scanning disk (for picture analysis) pierced by 24 apertures arranged at equal angular displacements along a spiral line. Each of the apertures has the shape of a picture element (pixel) and allows the light flux corresponding to the brightness of a picture element of the scene being televised to be incident, via a lens, onto a photoelectric cell. When the disk rotates, each aperture scans a given path (line) of the image field, and after one rotation of the disk the whole of the image has been scanned by 24 paths (lines). The process is then repeated, the number of scans of the complete image (rotations of the disk) per second being the frame rate. With a Nipkow disk the commencement of the line scanning, and also the frame scanning, is carried out automatically. Despite his patent, Nipkow never built a working prototype, as he had no way of amplifying the weak signal from the photocell.

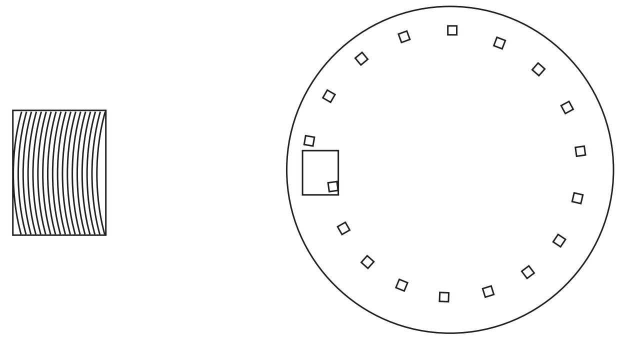


Figure 6. A Nipkow disk scanner and the paths of the apertures across the image plane.

John Logie Baird (1888–1946) employed various types of Nipkow disk type scanners in his early work on television and on 26 January 1926 demonstrated in London, for the first time anywhere, a rudimentary television system. Subsequently his basic scheme was adopted and adapted by many workers. Initially Baird used a "floodlight" method of illuminating his subjects (Figure 7(a)), but when this caused discomfort to his subjects he inverted the positions of the light sources and the photocells to produce a "spotlight" method of scanning (Figure 7(b)). Baird patented the method, as did Bell Telephone Laboratories (BTL) in New York, but unknown to Baird or BTL the method had been patented in 1910.

Both Baird and BTL adapted their systems to demonstrate color television, stereoscopic television (Baird), large screen television, multichannel television, and two-way television. BTL's experimental two-way system was in operation in New York for approximately one year from April 1930 and was used by around 17,000 persons. A novel application was observed when two deaf persons

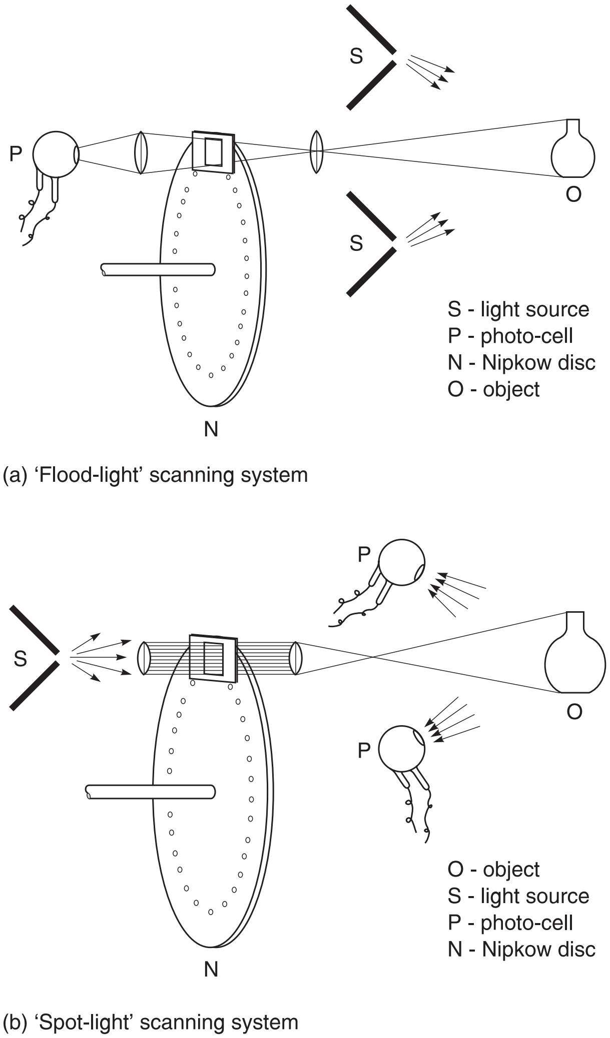


Figure 7. (a) Floodlight scanning system; (b) spotlight scanning system.

carried on a telephone conversation by reading each other's lips.

From 1880 to 1930 progress in television proceeded in an empirical manner. However, in a 1930 issue of the periodical *Fernsehen*, Moller and Kirschtein, in two separate papers, showed how the optical efficiencies of scanners could be calculated according to the principles of optics and photometry. Analyses of the relative efficiencies of the Nipkow disk and the Weiller mirror drum led to the conclusion that the aperture disk should be used at the transmitting end of a low-definition system of more than 40 lines or so. Below 40 lines it was advisable and more efficient to employ the mirror drum. Since Baird's low-definition system was based on 30 lines per picture, at 12.5 pictures per second, the British Broadcasting Corporation (BBC) utilized mirror drum scanners in its low-definition service. Further analysis showed that there was no optical advantage in using a Brillouin multilensed disk rather than a single lens-aperture disk combination for either floodlight or spotlight scanning. Again, analysis indicated that the Nipkow disk was unsuitable for image synthesis and from about 1932 its use declined. In some systems it was replaced by the mirror screw, as advocated by Hatzinger (1930). Essentially, this scanning element comprised a number of mirrors arranged like a spiral staircase on a central shaft so that the mirrors made one complete spiral. The mirror length was the same as the picture width the number of mirrors was equal to the number of scanned lines. For the 84-line images, first displayed in 1931, the mirror screw used 84 mirrors, each 100 millimeters long and 1 millimeter deep to produce a picture 84 by 10.0 millimeters in size. This image size was a substantial improvement over the display size achievable with the Nipkow disk.

By the early 1930s, the Radio Corporation of America (RCA), Farnsworth Television Inc., and Electric and Musical Industries (EMI) were undertaking much research and development work on the evolution of an all-electronic, high-definition television system. Their electronic cameras were

Figure 8. TV Dumont 180.

called the iconoscope, the image dissector, and the emitron respectively. Of these, the iconoscope and the emitron embraced the very important principle of charge storage. Philo Farnsworth's image dissector, basically, was the electronic equivalent of the Nipkow disk–photocell combination and was relatively insensitive compared to tubes of the iconscope and emitron type.

On 2 November 1936 the London BBC television station was inaugurated at Alexandra Palace. It operated with both electronic and mechanical scanners on an alternate basis. By 19 December 1936 it was apparent that the emitron cameras were greatly superior to those that used Nipkow disks. As a consequence, after 2 January 2 1937, only the 405-line, all-electronic system of EMI was operational.

See also **Iconoscope; Television: Color, Electromechanical; Television: Color, Electronic; Television, Beginning Ideas (Late nineteenth and Early twentieth Century)**

RUSSELL W. BURNS

Further Reading

Abramson, A. *The History of Television, 1880 to 1941*. McFarland, London, 1987.

Burns, R.W. *Television: An International History of the Formative Years*. Institution of Electrical Engineers, London, 1998.

Burns, R.W. *British Television: The Formative Years*. Institution of Electrical Engineers, London, 1986.

Burns, R. W. *John Logie Baird, Television Pioneer*. Institution of Electrical Engineers, London, 2001.

Television, Iconoscope, *see* **Iconoscope**

Television Recording, Disk

Recording television signals onto a disk has been a goal of engineers since the inception of television broadcasting, and in recent decades the effort to perfect a disk television recorder has resulted in many competing systems. The disk phonograph was established as the medium for home audio listening by the time television was tentatively introduced in the 1920s and 1930s, so it was natural that experimenters imagined a disk-based accompaniment to future home television sets. The earliest such system was probably that of Scottish inventor John Logie Baird. His "phonovision" disks, developed in 1926, recorded low-resolution television signals in a spiral groove on a phonograph disk. The output of the Baird camera, consisting of electrical pulses, was simply fed to an ordinary electromagnetic phonograph recording head, where the pulses modulated the movement of the cutting stylus.

Despite Baird's progress, the introduction of television in the U.S. and Europe in the late 1930s would be based on the live broadcast of programs, instead of the distribution of recorded programs as in the case of the phonograph. The invention of a successful videotape recording system by the Ampex Corporation (U.S.) in the 1950s led to predictions of home video recorders, and these began to appear in the 1960s (see Television Recording, Tape). The disk format was universally abandoned for commercial video recording, in part because the considerable bandwidth required for television signal recording demanded large areas of recording surface. This surface could be conveniently wound onto a spool in videotape systems, but would have required a very large disk to record even the 15- to 20-minute television shows of the day. However, for certain types of short-duration recording, the fast access possible only with a disk led to special-purpose devices such as the Ampex "instant replay" disk recorder, introduced in 1966 and widely used in the broadcast of sporting events. The Ampex instant replay recorder utilized thirty concentric tracks on a large magnetizable disk, each track representing a single frame of television. Thirty in-line heads read these tracks and allowed the broadcast of slow motion, fast motion, or still images of up to thirty seconds of video.

Improvements in technology also led to the revival of video recording on phonograph-like disk with the RCA Corporation's (U.S.) "Selectavision" system introduced in early 1981. Using the so-called capacitance electronic disk, Selectavision employed a very fine groove and stylus to achieve a recording time of over 60 minutes per side on 12-inch (300-millimeters) diameter plastic disk. The walls of the exceedingly fine groove interacted with a special stylus to establish a varying capacitance between the stylus and the disk. This varying capacitance contained the video information that was then processed and displayed on an ordinary television receiver. It was not possible for the consumer to make an original recording on the home players. Although considered a technical success, Selectavision was introduced at almost the same time as the soon-to-be popular Betamax and VHS home videotape systems. This competition, along with quality control problems and the lack of recording capability, caused RCA to discontinue Selectavision in 1986.

Meanwhile, various electronics manufacturers were experimenting with a video recording technology of a very different sort. These systems

utilized the relatively new technologies of digital recording and optical reproduction by laser. Telefunken (Western Germany), RCA, MCA (U.S.), Thomson (France), Sony (Japan), and Phillips (Netherlands) all demonstrated such videodisc systems by 1982, with the first commercial product being the Philips/MCA "DiscoVision," first sold in Atlanta, Georgia in December, 1978. The DiscoVision system recorded video information as pits on an aluminum-coated disk, and the player read the information using a reflected laser beam. However, it was not a digital recording, but rather a frequency-modulated recording system reminiscent of earlier videotape recording technology. DiscoVision failed almost immediately, as did most of the competing systems of the 1980s. A similar Pioneer "Laserdisc" format on 12-inch disks was one of the one or two videodisc formats to survive into the 1990s, although it sold in numbers that were dwarfed by VHS sales.

While the compact audio disc, the audio-only variation of this technology developed by Sony and Philips, proved to be a commercial success, they also provided the basis for a new type of laser videodisc technology employing digital data recording. Pioneer Corporation was among the first to offer a digital videodisc, and employed the same sort of pulse code modulation circuitry already in use in long-distance telephony, data recording on magnetic tape, and other applications. The shift to digital video recording was spurred by the introduction of the personal computer, which utterly changed the context of videodisc development. The CD-ROM, which became a popular way for software companies to deliver their products, was also once considered a competitor to VHS. While much smaller than the original digital videodiscs, it was in most other ways similar in operation. It was not common, however, to watch CD-ROM video presentations on home television sets, but rather on personal computers. However, some record companies began to include short video clips on a CD-ROM disk in 1991, calling this product the "interactive" CD, or CD-I. CD-I players were intended to be connected to both the home audio system and to a television receiver. There were several other similar formats during the 1990s, such as the Eastman Kodak Company's (U.S.) Photo CD, intended to compete with photographic prints for the storage of still images.

Video CD-ROMs achieved considerable commercial success, but did not have quite enough recording capacity to compete with videotape for the presentation of feature motion films, and at the peak of the CD-ROM's popularity it did not seriously threaten the videotape market. It was, however, a more flexible video format than VHS, capable of conveniently storing moving and still images and providing multiple grades of quality to suit the user's needs. Like all disk formats, the CD-ROM allows faster access to data located anywhere on the disk as compared to a linear tape format. Further, by 1996, a user-recordable disk called the CD-R was introduced, with a storage capacity comparable to a CD-ROM.

A refinement of the CD-ROM and CD-R is the Digital Video Disc (DVD, sometimes referred to as the Digital Versatile Disc) first offered in 1996. This format (which, like its predecessors, stores information as pits on a small laser-read disk) packs more information onto a CD-ROM size medium, and for the first time beats VHS tape in terms of both picture quality and recording time. Like the CD-ROM, it is equally applicable to both home video and personal computer applications. The DVD began to sell in large enough numbers by 1999 to garner both larger amounts of shelf space in video stores as well as inclusion as original equipment in many new personal computer systems. A short time later, recordable DVD drives became available for use in conjunction with personal computers. While the DVD may itself be superceded, it seems likely that the place of the laser-read disk format in television recording is assured for the near future.

See also **Audio Recording, Mechanical; Audio Recording, Tape; Television Recording, Tape**

David Morton

Further Reading

Burns, R. *John Logie Baird, Television Pioneer*. Institution of Electrical Engineers, London, 2000.

Daniel, E.D. *et al.* Eds. *Magnetic Recording: The First 100 Years*. IEEE Press, Piscataway, NJ, 1999.

Graham, M.B.W. *RCA and the Videodisc*. Cambridge University Press, Cambridge, 1986.

McLean, D.F. *Restoring Baird's Image*. IEEE Press, London, 2000.

Television Recording, Tape

The development of video tape recording is closely associated with the development of television. The rapid expansion of television broadcasting in the post-World War II period led to a demand for video recording to facilitate production of programming. Initially, this demand was satisfied by the use of motion picture camera to photograph images produced by a high-brightness cathode-ray

tube called a kinescope tube. Such kinescope recordings (the word came to describe a filmed recording although the word actually refers to the cathode-ray tube) were less than ideal, as the process introduces a variety of image defects due to poor resolution, compressed brightness range, nonlinearity, and film grain and video processing artifacts. Given that the resulting signal to noise ratio was often less than 40 decibels (dB), it is not surprising that television producers sought an alternative recording method to make a permanent document for rebroadcast or archiving.

Initial experiments with video recording in the early 1950s were based on magnetic tape technology developed for audio tape recording, a technology that had reached maturity in the late 1940s. However, video recording proved to be a much more difficult technical challenge than audio recording, due to the higher bandwidth required. In contrast to audio signals, which range from about 20 hertz to 20 kilohertz, or roughly 10 octaves, video signals require a recording range from about 30 hertz up to 5 megahertz, more than 17 octaves. With contemporary technology, this required very high tape speeds, which consumed enormous amounts tape. Moreover, these high speeds led to unacceptable wear on tape recording and reproducing heads.

Early experiments, which used a fixed recording head, attempted to solve the problem by using multiple tracks on a single piece of tape, either by running the tape past the head back and forth, or electronically splitting the incoming signal into several frequency bands and recording on separate parallel tracks on the tape. Attempts to solve the mechanical and electronic issues associated with these systems were unsuccessful, and no such machine was ever commercialized.

Led by Charles Ginsburg, the Ampex Corporation solved the problem of video recording by adopting a new type of recording head. The VRX-1000, introduced in 1956, was the first practical videotape recorder. At $50,000, it was used by television networks for videotape delayed broadcast, rather than home recording. Referred to as the quadraplex system, the Ampex machine moved tape past a thin wheel on which were mounted four recording heads that rotated at right angles to the tape motion. This system allowed writing on thin parallel tracks on the tape at very high effective speeds, thereby allowing the capture of television pictures. Initial quadraplex machines were capable of monochrome recording and subsequent improvements allowed the recording of color video. Quadraplex machines were used until the mid-1970s, when they were supplanted by helical scan recorders.

Quadraplex video recorders were in widespread use in television production by the late 1950s. However, these machines were expensive and required constant adjustment to perform well, and manufacturers sought to develop less expensive machines based on emerging transistor technology. In 1958 Ampex developed the helical scan recorder, which wrapped the recording tape in a spiral path around a rotating cylindrical recording head. This allowed the recording of a complete TV field on a single track, and considerably simplified the electronics needed to process the image. However, due to problems with timebase correction, helical scan systems did not match the performance of quadraplex recorders until the early 1970s. After that time, advances in large-scale integration of silicon devices improved the quality and lowered the cost of helical scan recorders so that they replaced quadraplex machines for studio use by the late 1970s.

Helical scan machines were marketed and used in a number of other applications during the 1960s. Television stations in smaller markets or that were not affiliated with the large networks were willing to use helical scan machines with lower levels of performance due to their lower cost. Portable helical scan machines were also developed during the 1960s allowing taping of news and sports recording, though these units were rather heavy and required carts to move them.

Continued development led to the first successful video cassette recorder, the U-Matic system, jointly developed by the Japanese firms Sony, Matsushita, and JVC, and introduced in 1969. The tape cassettes were expensive (almost $100 for one hour), but the machines were widely adopted by institutional users to make training films, replacing 16-millimeter film. In 1974 Sony developed a portable camera system for CBS, which was very successful, but still heavy—one person was needed to carry the recorder, the other the camera. Subsequent competition led to the introduction of smaller recording camera units.

The first home video recorders were introduced by Sony and Matsushita in the mid-1960s, but these reel-to-reel machines were very expensive, had poor picture quality and could only record for a short period of time (less than one hour). With the introduction of chromium dioxide tape in the late 1960s, higher recording densities became possible, and home machines based on the U-Matic format were introduced. Their high cost largely limited their use to institutional settings,

and it was not until the introduction of the VHS and Beta cassette formats in the mid-1970s that prices declined to a level that led to widespread consumer acceptance. The key technical innovation used in both VHS and Beta was azimuth recording, which allowed tracks to be recorded much closer together, reducing the amount of tape needed for a given recording time. VHS and Beta machines offered similar levels of performance, but the earlier introduction of longer recording times by VHS machine makers led to their eventual domination of the market by the late 1980s. Subsequently, smaller size cassettes, such as the 8 mm and VHS-C formats, were developed for use in portable video cameras for consumer use.

Although digital recording is an idea that dates to the late 1930s, the digital recording of video signals was delayed for many years by the difficulty of developing adequate analog to digital converters. The rise of large-scale integrated circuits in the 1970s solved this problem, and the first digital video recorders were operating in laboratories by the mid-1970s. Negotiations over a common digital recording format extended into the early 1980s, and it was not until 1986 that the first generation of digital video recorders were marketed by Sony of Japan and Bosch-Fernseh of Germany. By the end of the twentieth century, digital video recording dominated broadcast recording.

See also **Audio Recording, Tape; Television Recording, Disk**

MARK H. CLARK

Further Reading

Daniel, E.D., Mee, C.D. and Clark, M.H. *Magnetic Recording: the First 100 Years*. IEEE Press, 1999.

Daniel, E.D. and Mee, C.D. *Magnetic Storage Technology*, 2nd edn. McGraw-Hill, 1996.

Marlow, E. and Secunda, E. *Shifting Time and Space: the Story of Videotape*. Praeger, New York, 1991.

Nmungwun, A.F. *Video Recording Technology: Its Impact on Media and Home Entertainment*. L. Erlbaum Associates, Hillsdale, NJ, 1989.

Thin Film Materials and Technology

Thin film technology—the growth or deposition of mechanical strengthening, optical, electronic, magnetic, or semiconductor materials in an ultrathin layer—resulted from rapid development of materials science and technology in the late twentieth century. Thin film technology aided the development of devices such as transistors and microelectronic components such as diodes, capacitors, sensors, and resistors, microelectromechanical systems (MEMs), and solar cells. Thin films enable transparent conductive coatings for touch screens and thin conductive films on magnetic read–write heads that allow increased magnetic data storage capacity. In aerospace engineering thin films contribute to strengthening against wear friction and corrosion. They are also used to coat microcircuits and optical lenses to withstand stress and extreme temperatures, protect them from damage and wear, give antireflection or polarizing coatings, and improve durability and performance.

Chemists create thin films, usually only a few micrometers or less thick but potentially nanometers for monolayers, through some method of depositing atoms from a source material target onto a foundation called a substrate (often a silicon wafer, but could be metal or glass). Deposition techniques and material sources can be adjusted to design thin films with tailored properties or thickness to meet specific industrial needs and conditions.

Thin films technology is based on research and processes related to vacuum, gas diffusion, and thin material layers that had gradually developed in the nineteenth century. In 1852, Sir William Robert Grove observed in experiments involving electric discharges between electrodes in a low-pressure atmosphere that metal from one of the electrodes was deposited on the glass walls containing the electrodes and the gas. This "vapor deposition" of metal films, also reported by Michael Faraday in 1854 and Julius Plücker in 1858, is now known to be caused by sputtering (see below). By 1887, Robert Nahrwold heated platinum wires inside a vacuum to deposit material for thin films.

The first commercial use of thin films was in 1901, when mass production of Edison's "gold molded" cylinder phonograph records was enabled by a vacuum coating process that deposited gold vapor from gold electrodes. With awareness of possible industrial uses for vacuums, research into vacuum equipment accelerated, and then shifted to applications by the mid-twentieth century. Innovation of existing thin film processes increased in the mid-twentieth century, quickly escalating from the 1960s through the turn of the century, with chemists adapting processes and materials to fabricate new thin films to create desired structures. Developments paralleled microelectronics demands, particularly the need for transistors and integrated circuits to have pure layers with no microstructural defects that could affect electronic properties. Thin film technology also progressed as processes and underlying knowledge of materials

such as electroceramics advanced. For example, discovery of high-temperature superconducting oxides (ceramics) in the late 1980s stimulated development of new thin film deposition techniques, owing to potential applications in superconducting electronics.

Technologists consider vacuum evaporation to be the most efficient and productive thin film deposition method. Physical vapor deposition (PVD) processes usually involve depositing material from a vaporized solid or liquid target source by moving atoms through a vacuum or low-pressure gas or plasma to condense onto a substrate.

Vacuum process technology relies on vacuums that are as empty as possible of any particles and gases that might interfere with materials being deposited within the vacuum. Technologists heat a selected source material in a vacuum so that sublimation or evaporation, by thermal or electron beam heating, results in atoms or molecules forming a film on a substrate. Substrate materials are often metals or glass composed of aluminum, silicon, or beryllium that are smooth, mechanically strong, chemically stable, and have desired thermal qualities. The quality of thin films is diminished with exposure to any moisture or contaminants that enter the vacuum if seals leak, pressure is not maintained, or the chamber is not cleaned.

Molecular beam epitaxy (MBE) methods in which the film crystal structure is "ordered" as it is deposited (growth of the deposited crystal is oriented by the lattice structure of the substrate) were developed by several chemists in the U.S. (including Alfred Cho and John Arthur at Bell Labs), Europe, and Asia during the 1960s and 1970s. Vapors from heated sources form beams that condense onto a heated substrate to create thin crystalline films just as in vacuum evaporation. However the timing and content of these beams can be carefully controlled by shutters in front of the heated sources that can close and block the atomic beams. In the 1960s, this assisted the subsequent development of integrated circuitry. The molecules are placed one layer at a time on the substrate, producing monolayer films that are more suitable for tiny electronics applications than films in which the substrate and polycrystalline film are not so evenly matched and the film's grain size is larger than the circuits for which the films will be used. This process occurs in an ultrahigh vacuum compared to other thin film deposition techniques, and as a result, films tend to be cleaner than those created by other methods. Because MBE relies on vacuum pressures and hygienic measures often too unstable to sustain, technologists devised technology in which various aspects of MBE deposition occur separately in a series of connecting chambers which process crystalline wafer substrates. After being decontaminated and prepared, a wafer moves on a platform along a track and is heated from 500 to 900°C before entering the growth chamber where films are formed.

Sputtering deposition differs from high-vacuum techniques because a rare gas such as argon is always moving in the vacuum chamber. Gas flow and a throttle valve determine gas pressure in the chamber. The rare gas is ionized in the electrical field formed by the substrate and diode or magnetron target in the chamber, creating an ionized plasma with an overall neutral charge. The accelerated ions strike the target causing its atoms to be ablated, and the target atoms are directed towards the substrate. Technologists often choose sputtering to create thin films from refractory sources including tungsten because deposition can occur at less than the materials' normal melting point. Sputtering also appeals because technologists can design films similar to alloy materials or compound sources. This method efficiently shifts atoms from targets to substrates to produce a film almost identical to its source. Evaporation deposition techniques are not as consistent because fluctuating vapor pressure affects how source material is deposited.

Unlike PVD methods where material is removed from a solid target, chemical vapor deposition (CVD) uses reactive carrier gases to form new material on a heated substrate. These gases either break down into reactive precursors or interact, and the resulting materials coat substrates. CVD offers chemists the capability to create a large variety of thin films suitable for numerous uses. CVD technologists use a reactor that provides the energy necessary to cause chemical reactions to deposit material to form films. They have innovated processes for specific material sources such as metal-organic CVD for organometallic materials used in semiconductors. Plasma is the energy source for reactions in plasma-enhanced CVD, in which vaporized compounds in the plasma trigger the coating of substrates. A vacuumless process, atmospheric pressure CVD moves substrates on a belt through the chamber to deposit materials for films used to coat silicate glasses. In this technique, coating colors can be achieved by the selection of specific compound sources. Diamond thin films that increase the surface hardness of cutting tools have been a popular product of CVD methods. Based on work patented by American William G.

Eversole in the 1950s and developed in England by John C. Angus, H.A. Will, and W.S. Stanko in the next decade, CVD methods have continued to improve throughout the late twentieth century for optical and electronic usages.

Ion beam deposition (IBD) refers to several interconnected processes that create predictable quality and characteristics in thin films. Kasturi L. Chopra and M. R. Randlett first used ion beam sputtering to make thin films in the mid-1960s. Chemists appropriated techniques for IBD, using ion beams that are usually broad and high energy to prepare substrates prior to deposition by removing contaminants. They then utilize ion beams to deposit materials by sputtering techniques as described above or as an aid for other methods to achieve optimal film production. Typically, the ion beam is aimed at a metal. As a result, target material sputters onto substrates, forming thin films. Often, an additional ion source known as the ion assist source (IAD) provides energy in ions that thicken and stabilize the films' surfaces and enhance their strength and capabilities. Engineers have appropriated IBD to reinforce magnetic heads with carbon films.

See also **Ceramic Materials, Coatings, Pigments and Paints; Crystals, Synthetic; Integrated Circuits, Fabrication, Materials and Industrial Processes, Nanotechnology, Materials and Applications, Optical Materials, Semiconductors, Crystal Growing and Purification; Transistors**

ELIZABETH D. SCHAFER

Further Reading

Elshabini-Riad, A.A.R., and Barlow III, F.D. *Thin Film Technology Handbook*. McGraw-Hill, New York, 1998.

Francombe, M.H. and Vossen, J.L., Eds. *Thin Films for Emerging Applications*. Academic Press, Boston, 1992.

George, J. *Preparation of Thin Films*. Marcel Dekker, New York, 1992.

Hopwood, J.A., Ed. *Ionized Physical Vapor Deposition*. Academic Press, San Diego, 2000.

Mahan, J.E. Physical Vapor Deposition of Thin Films. Wiley, New York, 2000.

Mattox, D. *Foundations of Vacuum Coating Technology*. Noyes, Norwich, NY, 2003.

Ohring, M. *The Materials Science of Thin Films*. Academic Press, Boston, 1992.

Ramesh, R. *Thin Film Ferroelectric Materials and Devices*. Kluwer, Boston, 1997.

Schuegraf, K.K., Ed. *Handbook of Thin Film Deposition Processes and Techniques: Principles, Methods, Equipment, and Applications*. Noyes, Park Ridge, NJ, 1988.

Smith, D.L. *Thin-Film Deposition: Principles and Practice*. McGraw-Hill, New York, 1995.

Venables, J. *Introduction to Surface and Thin Film Processes*. Cambridge University Press, Cambridge, 2000.

Wasa, K. and Hayakawa, S. *Handbook of Sputter Deposition Technology*. Noyes, Park Ridge, NJ, 1992.

Timber Engineering

Timber engineering is the technology of creating wood products not found in nature. Manufactured lumber has characteristics superior to those found in its individual components.

Glued layers of hardwoods or veneers were used for decoration by the ancient Egyptians. The first plywood made from layers of softwood was developed in the early twentieth century. In 1905, the directors of Portland, Oregon's Lewis & Clark Exposition asked the Portland Manufacturing Company to devise for display some new and unusual wood product. To bring attention to the region's rich timber resources, the company manufactured the first Douglas fir plywood.

Appeal for the product was immediate and worldwide in scope. Mills everywhere produced thin rectangular sheets of the lightweight wood product. Assembled so that the grain of each ply alternated direction by 90 degrees, it was strong, warp resistant, dimensionally stable, and did not split. It was useful in such applications as door panels, drawer bottoms, crates, trunks, and partitions. If the material had one shortcoming, it was the tendency to delaminate when exposed to dampness. Adhesives were not waterproof and early plywood was limited to interior or protected exterior use.

Contemporary to the development of softwood plywood was the refinement of glued laminated lumber or glulam technology. The process, which allowed the use of thinner (younger) trees, produced beams, columns, and lumber to serve as structural elements. Although glued timbers were first used in Britain during the 1860s, the technology developed no further. A building boom in Germany around 1900 and a shortage of large size timbers prompted lumber mill owner Otto Hetzer of Weimar, to investigate methods for fabricating long beams from the materials at hand. Hetzer eventually received five patents for methods of assembling or laminating wood segments into beams of great length.

The most practical system was a laminate of individual narrow planks no more than 5 centimeters thick. Beams were made thicker and wider by adding planks. Assembled with their grain running parallel, the slender "lams" were relatively

pliable and could be bent into job-specific shapes during gluing. Although no metal fasteners were used to join the lams, finished lumber had great strength as well as length. Thus, glulam lumber was useful in extended clear span structures.

The first public demonstration of the technique was at the Brussels World's Fair in 1910. The technology spread quickly throughout Germany and was later adopted in Switzerland, Sweden, the Netherlands, and Italy. Although the U.S. investigated the system in 1920, it was 1934 before the first laminated lumber structure was erected there. Thereafter, the technology was embraced for public structures nationwide. Glulam had much in its favor: it demonstrated an economic use of materials and made possible new innovative building designs; prefabricated components meant faster assembly and cost saving. Over the years the technology changed little.

Plywood's inherent strength and lightness lent itself well to the construction of aircraft, which were made primarily of wood. In 1915 the LWF (laminated wood fuselage) Engineering Company of New York City fabricated one of the first molded wood aircraft fuselages. Aircraft builders experimented with plywood for airframes, wings, and tails throughout the 1920s and 1930s.

Adhesives, critical to the growing engineered lumber industry, varied little from those used in previous centuries. During the 1920s, glues were still derived from animal and vegetable matter. Bone and hide wastes, casein from cow's milk, and in 1926 soy beans were among the materials used to produce glue.

The 1930s brought several important breakthroughs in timber engineering technology. Heat along with pressure became part of the adhesive curing process. A significant turning point was reached with the introduction of two resin adhesives made from synthesized formaldehyde gas. Water-resistant urea formaldehyde and waterproof phenol formaldehyde adhesives dramatically increased the potential uses for engineered lumber.

Worldwide shortages of vital metals during World War II placed plywood in an important position in the military industries of several nations. Undoubtedly, the best known and most successful aircraft was the British Mosquito bomber. Designed with twin engines and capable of speeds of over 600 kilometers per hour, the fuselage was assembled from molded plywood halves. Plywood was used by the Allies in their rush to produce naval vessels. The coastal forces of Britain's Royal Navy employed a fleet of plywood motor-torpedo and motor-gun boats (MTB and MGB). American-built Patrol Torpedo (PT) boats while not made entirely of plywood, used much in their construction. Each steel-hulled American-built Liberty ship contained 2325 square meters of plywood. Landing craft used to deliver vehicles and personnel to the beaches of Normandy were part of a fleet of 20,000 such plywood vessels made by the Higgins Company of the U.S. The Axis powers used plywood as well, perhaps most notably in Japan's lightweight explosive laden plywood suicide boats.

Since the end of World War II plywood's use rose dramatically, primarily in construction. Plywood subflooring, decking, siding, and roofing revolutionized the construction of light frame buildings, especially in the U.S.

In the postwar era, particle board also became a component of the construction industry. Particle board was first produced early in the century, but its commercial production dates only from the 1940s. There was a definite economic benefit in boards made of wood chips, shavings, and trimmings as well as agricultural fibers. Bonded with water-resistant adhesive it was used most often for interior work.

The shortage of large timber first observed at the beginning of the century was even more acute at the end. What was formerly considered waste was turned into reliable, highly predictable dimensioned lumber of uniform composition.

Several new engineered lumbers appeared late in the century. Panels of oriented strand board (OSB) were manufactured from rectangularly shaped wood strands arranged in layers with alternating grain. Manufactured like plywood, they had common characteristics. Chipboard panels were yet another lumber made from what was a formerly waste. Particles of resin-coated softwood were bonded together by heat and high pressure. Because it was unable to tolerate dampness, it too was used primarily for interior work.

During the 1970s, new applications were devised for laminated veneer lumber (LVL). Made from plys or veneers, the laminations are glued and processed to form material with the dimensions of sawn lumber. While built up of thin plys as in plywood, it differs in that all grains are aligned. The segments can be assembled to produce material whose finished length is limited only by the size of the machines used to manufacture it and the means of transporting it.

With ever-dwindling supplies and an expanding market for lumber products, new and ingenious methods continue to be devised to utilize available wood to its fullest extent.

See also **Adhesives; Buildings, Prefabricated**

William E. Worthington, Jr.

Further Reading

Andrews, H.J. *An Introduction to Timber Engineering*. Pergamon Press, Oxford, 1967.

Baldwin, R.F. *Plywood Manufacturing Practices*. Miller Freeman Publications, 1975.

Bulkley Jr., R.J. *At Close Quarters: PT Boats in the United States Navy*. U.S. Government Printing Office, Washington D.C., 1962.

Cour, R.M. *The Plywood Age, A History of the Fir Plywood Industry's First Fifty Years*. Binsford & Mort, Portland, 1955.

Levin, E. Ed. *Wood in Building*. Architectural Press, London, 1971.

Mondey, D., Ed. *The Hamlyn Concise Guide to British Aircraft of World War II*. Hamlyn Aerospace, London, 1982.

Muller, C. *Holzleimbau, Laminated Timber Construction*. Birkhauser, Basel, 2000.

Perkins, N.S. *Plywood: Properties, Design, and Construction*. Douglas Fir Plywood Association, Tacoma, 1962.

Sellers Jr., T. *Plywood and Adhesive Technology*. Marcel Dekker, New York, 1985.

Tissue Culturing

The technique of tissue or cell culture, which relates to the growth of tissue or cells within a laboratory setting, underlies a phenomenal proportion of biomedical research. Though it has roots in the late nineteenth century, when numerous scientists tried to grow samples in alien environments, cell culture is credited as truly beginning with the first concrete evidence of successful growth *in vitro*, demonstrated by Johns Hopkins University embryologist Ross Harrison in 1907. Harrison took sections of spinal cord from a frog embryo, placed them on a glass cover slip and bathed the tissue in a nutrient media. The results of the experiment were startling—for the first time scientists visualized actual nerve growth as it would happen in a living organism—and many other scientists across the U.S. and Europe took up culture techniques. Rather unwittingly, for he was merely trying to settle a professional dispute regarding the origin of nerve fibers, Harrison fashioned a research tool that has since been designated by many as the greatest advance in medical science since the invention of the microscope.

Indeed, after this initial experiment, Harrison left it to others to take up the mantle of culturing tissue; and it is Alexis Carrel who became inextricably associated with the technique in its early history. Carrel, a French surgeon, worked at the Rockefeller Institute in New York, where he cultured *in vitro* the cells of many warm-blooded animals. Generally, from 1910 until 1920, research into tissue culture consisted primarily of simply determining how many tissues could survive in cultures and for how long. In 1912, Carrel claimed to have established an immortal strain of chick cells—a claim that attracted a huge amount of international press attention and alerted the public to the potential of this technique for the first time.

By contrast, many scientists continued to lament both the lack of medically relevant research derived from cells in culture and the difficulty of getting cultures to survive for any length of time *in vitro*; for many years opinion remained divided on the method's relevance and prospects. However, a number of technical advancements in the 1940s began to sway scientific opinion. A major breakthrough occurred in 1943 when Wilton Earle, working at the National Institutes of Health, developed the first permanent mammalian "cell line." This term applies to a culture of cells that either through innate or induced malignancy does not die *in vitro*, enabling scientists to work with a standardized model that can be easily distributed. After World War II widespread use of antibiotics prevented the hitherto disastrous problem of continual bacterial infection and made cell survival in culture far less difficult to attain. No less important were the development of optimal, artificial growth media in the U.S. in the early 1950s and improved methods such as nitrogen freezing for long-term storage of cultures.

Subsequently, cell culturing was used in researching most medical problems, replacing the traditional animal models in experiments to determine the cellular basis of cancer. The International Tissue Culture Association was founded in 1946, which aimed to foster an international community to disseminate findings and also train those lab workers who now chose to work on cultures. Thus, at the beginning of the 1950s, scientific unanimity had been reached on the technique's value.

In 1951, researchers at Johns Hopkins University in Baltimore, Maryland, achieved the first human cell line—HeLa—derived from the cervical tumor of an American woman named Henrietta Lacks. This cell line was very widely used, but neither Lacks nor her family learned of HeLa until well after her death, a distressing episode that the Clinton administration formally acknowledged in 1997. As well as their use in research on cancer, cultures began to be inoculated with viruses to produce vaccines against human diseases. Indeed, HeLa cells became the first to be

produced on an industrial scale at Tuskegee, Alabama, for the purposes of producing polio antibodies. Such was the demand for mass-produced HeLa that Tuskegee scientists distributed over 600,000 cultures within two years. Other researchers followed suit, and cell cultures rapidly became the experimental model of choice for virologists. However, HeLa possessed a remarkable propensity to contaminate and colonize other cultures, which ultimately undermined millions of dollars of research on what transpired to be falsely labeled HeLa cells, including a much vaunted collaborative project between the U.S. and USSR in the 1960s as part of Richard Nixon's so-called war on cancer.

In 1962, a centralized store of cell cultures was established at the American Type Culture Collection in Washington in order to ensure ready access to a wide variety of cultures with (crucially, in the light of the HeLa fiasco) defined and documented characteristics. In 1975, the value of cultures increased further, with the production of specialized hybridoma cells by Milstein and Köhler, in Cambridge, England. These cells, the product of fusion between a cancer cell in culture and an antibody-producing lymphocyte cell, produce high quantities of a specified antibody—a process that is simultaneously of immense scientific and commercial value. Indeed, in some cases it has been estimated that the products secreted by certain cells in culture can be worth up to $3 billion, so it is little wonder that commercial firms value the technique as highly as academic researchers.

From the 1980s, cell culture has once again been brought to the forefront of cancer research in the isolation and identification of numerous cancer causing oncogenes. In addition, cell culturing continues to play a crucial role in fields such as cytology, embryology, radiology, and molecular genetics. In the future, its relevance to direct clinical treatment might be further increased by the growth in culture of stem cells and tissue replacement therapies that can be tailored for a particular individual. Indeed, as cell culture approaches its centenary, it appears that its importance to scientific, medical, and commercial research the world over will only increase in the twenty-first century.

DUNCAN L. WILSON

Further Reading

Brown, R.M. and Henderson, J.H. M. The mass production and distribution of HeLa cells at Tuskegee Institute, 1953–1955. *J. Hist. Med. Allied Sci.*, 38, 415–431, 1983.

Freshney, I.R. *Culture of Animal Cells: A Manual of Basic Technique*. Alan R. Liss, New York, 1983.

Gold, M. *A Conspiracy of Cells: One Woman's Immortal Legacy and the Medical Scandal It Caused*. State University of New York Press, Albany, 1986.

Landecker, H. Immortality *in vitro*: A history of the HeLa cell line, in *Biotechnology and Culture: Bodies, Anxieties, Ethics*, Brodwin, P., Ed. Indiana University Press, Bloomington, IA, 2000.

Landecker, H. New times for biology: nerve cultures and the advent of cellular life *in vitro*. *Stud. Hist. Phil. Boil. Biomed. Sci.*, 33C, 4, 667–694, 2002.

Masters, J.R. W. Human cancer cell lines: fact and fantasy. *Nat. Rev.: Mol. Biol.*, 1, 3 233–236, 2000.

Witkowski, J.A. Dr. Carrel's immortal cells. *Med. Hist.*, 24, 129–142, 1980.

Useful Websites

American Type Culture Collection: http://www.atcc.org

Tomography in Medicine

The word tomography derives from the Greek word *tomos* meaning a section, and the technique was proposed by Gustave Grossman in papers published in 1935 describing the experimental details of tomography and its clinical application. A tomographic image is essentially a picture of a slice of the patient's anatomy. Developments in x-ray tomography culminated at the end of the twentieth century with the invention of computed tomography (CT), which has found numerous applications in diagnostic medicine including scanning for cancer, subdural hematomas, ruptured disks and aneurysms. It substantially reduced the requirement for exploratory surgery. Tomographic techniques are also employed in other diagnostic imaging modalities such as ultrasonography. Early B-Mode ultrasound scanners were mounted a single transducer element on an articulated arm to produce an acoustic tomographic slice of the body. The arm was moved through a series of positions to allow the "slice" of data to be collected.

In standard x-ray imaging, three-dimensional patient information is projected onto a two-dimensional surface. Each point on the image is a projection of the attenuating properties of all the tissues along the x-ray beam's path. Tomographic techniques sought to examine a section through the body ignoring the tissue above and below the section of interest. In early x-ray tomography, the x-ray tube traveled in an arc above the patient while the film cassette in the table, beneath the patient, traveled in the opposite direction so that x-rays remain incident on the film throughout the exposure (Figure 9). The patient was positioned so

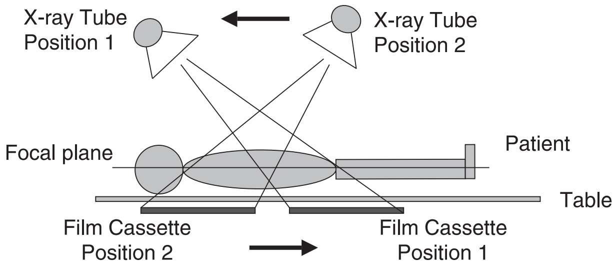


Figure 9. Motion of the x-ray tube and film cassette in x-ray tomography.

that the area of interest within the body was at the focal plane of the x-ray tube. When the x-ray tube moved through the tomographic angle, objects above and below the focal plane were blurred, thereby largely obliterating their contrast. As a result, the objects at the focal plane were seen more clearly than objects above and below the plane. The tomographic angle determines the slice (depth of objects in the focal plane) thickness. As the out-of-plane anatomy is not removed from the image, but blurred out, the effective image contrast is reduced. The clinical importance of tomography was quickly established, and the first atlas of clinical tomograms was compiled in 1940.

The development of CT followed from the work of numerous scientists who had built a variety of apparatus that aimed at true section imaging, which is achieved when the presence of planes above and below the image are of no consequence. Gabriel Frank (1940) showed how a transverse axial section can be sharply imaged without contribution from adjacent sections. Images were reconstructed by a form of optical back-projection, a method that emulates the image acquisition sequence in reverse. This was a form of CT without the computing, three decades before CT was developed. A Japanese radiologist, Shinji Takahashi, conducted a series of experiments in the 1940s and 1950s which also anticipated the later developments in CT. W.H. Oldendorf (1961) designed an apparatus to measure radiodensity at a point, uncoupling the effects of all other points in the same plane. This was achieved by arranging that the signals were at different frequencies. Oldendorf is often credited with producing the first laboratory CT of a "head" phantom.

Development of emission tomography methods overlapped with the development of transmission CT. Tomographic nuclear medicine was studied in the 1960s and developed into computerized tomography with the invention of single photon emission computed tomography (SPECT) in the 1970s. Experimental work leading to the development of modern positron emission tomography (PET) systems began in the 1950s.

In 1972 Godfrey Hounsfield, a British engineer employed by Electric and Musical Industries (EMI), developed the first CT system. CT technology quickly evolved through a series of generations with most of the initial developments focusing on increasing the scanning speed and thus reducing problems with patient motion. Developments in computing complemented these advances enabling rapid reconstruction of images from the huge data sets obtained. Hounsfield's scanners included what have become known as first, second, and third generation scanners developed between 1972 and 1977. Although numerous designs of scanners were patented (particularly in later generation scanners) the following broadly applies to the designs.

First generation systems used an x-ray tube, which emitted pencil beam x-rays. Transmitted x-rays were measured by a single detector (per slice). The tube and detector were translated across the object to acquire x-ray transmission data (Figure 10). The detector and x-ray tube were then tilted at an angle of 1 degree, and the translation process was repeated until the system had completed a 180-degree rotation and a complete set of CT data was obtained. Approximately 30,000 projections were taken with scanning times running to several minutes per slice.

Second generation CT scanners employed a linear array of 30 detectors to collect x-ray transmission data. The x-ray beam was collimated to a narrow fan. Second generation scanners did not have as good scatter rejection as first generation but slice acquisition time was as much as 15 times faster. Additional data detection capability was used to increase speed of acquisition and image quality.

By the mid-1970s third generation scanners had been developed, and these were the first fast medical scanners. More than 800 detectors were employed, arranged in an arc encompassing the width of the patient. This eliminated the need for translation motion. In this design both tube and detectors rotate together inside a gantry to acquire the necessary projections. Scan times per slice were reduced to approximately 5 seconds.

In the late 1970s, fourth generation CT scanners were built with approximately 5000 individual detectors arranged in a 360-degree ring around the patient (Figure 11). The detectors remain stationary while the tube is rotated.

Obtaining CT images requires taking x-ray transmission measurements at numerous angles

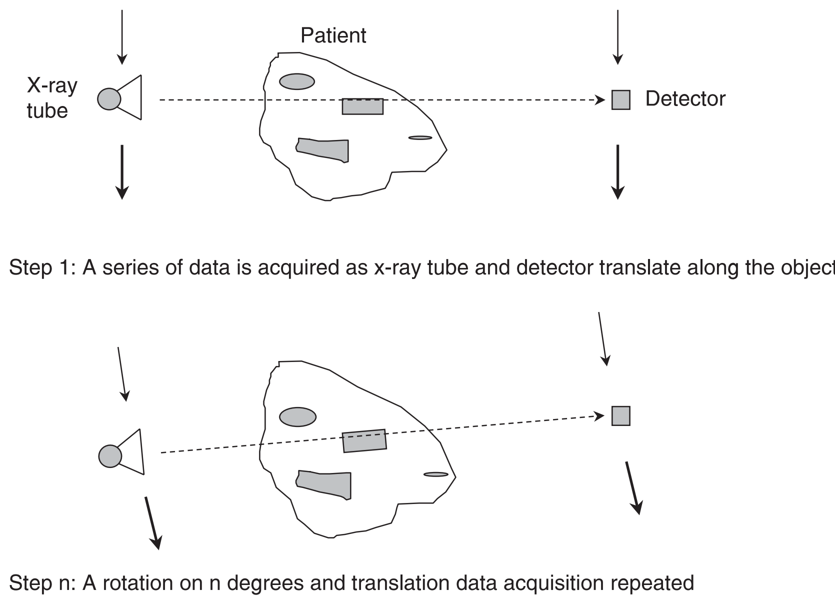


Figure 10. First generation computed tomography (CT).

around the object. This is roughly analogous to precisely locating the seeds inside an orange by probing it from many points with a thin wire. The image may be built up through a back-projection technique, which is essentially a mathematical method for establishing the geometry of the object based on data collected from x-rays taken at multiple angles through the object.

The mathematical principles underlying CT were developed by an Austrian mathematician, Johann Radon, in 1917. Radon theoretically demonstrated that an image of an unknown object could be reproduced from an infinite number of projections through the object. Images are reconstructed from the raw data through sophisticated algorithms. This involves processing 800,000 or so projections for each picture element (pixel) displayed, as well as applying image-processing algorithms. Filtered back projection is a common algorithmic image reconstruction technique. The process emulates the acquisition procedure in reverse and convolves the solution with a filter of kernel to correct for known errors in the reconstruction process and to emphasize particular characteristics in the image. Different kernels are available for types of soft tissue and bone and may be selected by the operator.

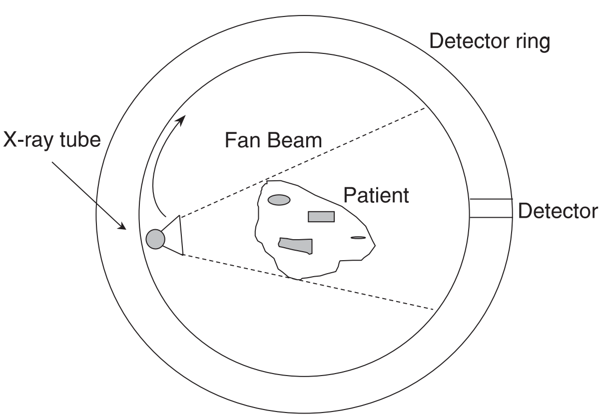


Figure 11. Fourth generation scanner. The detectors are fixed in a ring around the patient, and the x-ray tube rotates during acquisition.

Digital technology for display of images became prominent in the 1980s. Image displays for CT typically have a 512 by 512 display matrix with each pixel in the matrix having a distinct value (or CT number) based on the calculated attenuation value for the corresponding volume element (or voxel) in the patient. Images are usually presented as a series of two-dimensional images on a monitor, with each image representing a slice through the patient. Computer programs can also produce three-dimensional reconstruction of CT images.

The development of CT has been compared in significance to Wilhelm Roentgen's discovery of x-rays. Hounsfield won the Nobel Prize for Medicine rather than Physics, reflecting the considerable clinical benefits in producing high-contrast images of slices through a patient. Developments in CT technology through the 1980s included the helical CT scanners that acquire information while the table moves the patient through the ring of detectors. The effect of this linear translation through a rotating acquisition is to produce a helical or spiral motion of the x-ray tube around

the patient. Helical scanners increase the efficiency of multislice acquisition and reduce overall scanning time.

The 1990s saw the development of multidetector arrays. In these systems slice width is determined by the way the detectors are grouped together rather than by beam collimation, as was the case in earlier generations. Multidetector arrays bring CT imaging systems closer to acquiring a full three-dimensional data set by reducing slice thickness to about 0.5 millimeters.

See also **Positron Emission Tomography (PET); Ultrasonography in Medicine; X-Rays in Diagnostic Medicine**

Colin Walsh

Further Reading

Bushberg, J., Seibert, J.A., Leidholdt, E.M. and Boone, J.M. *The Essential Physics of Medical Imaging*, 2nd edn. Lippincott Williams & Wilkins, Philadelphia, 2002.

Curry, T.S., Dowdey, J.E. and Murry, R.C. *Christensen's Physics of Diagnostic Radiology*. 4th edn. Lea & Febiger, 1990.

Goodenough, D.J. Tomographic imaging, in *Handbook of Medical Imaging, Volume 1. Physics and Psychophysics*, Beutel, J., Kundel, H.L. and Van Metter, R.L., Eds. SPIE Press, 2000.

Grossman, G. Lung tomography. *Brit. J. Radiol.*, 8, 733, 1935.

Hounsfield, G. Computerized transverse axial scanning (tomography), part 1: description of system. *Brit. J. Radiol.*, 46, 1023, 1973.

Mould, R.F. *A Century of X-Rays and Radioactivity in Medicine*. Institute of Physics Publishing, Bristol, 1993.

Swindell, S. and Webb S. X-ray transmission computed tomography, in *The Physics of Medical Imaging*, Webb S., Ed. Institute of Physics Publishing, Bristol, 1988.

Webb, S. *From the Watching of Shadows: The Origins of Radiological Tomography*. Institute of Physics Publishing, Bristol, 1990.

Wegener, H.O., *Whole Body Computed Tomography*, 2nd edn. Blackwell Scientific, 1993.

Useful Websites

http://www.nobel.se/medicine/laureates/1979/hounsfield-autobio.html

Transistors

In 1906, the American inventor Lee de Forest developed a triode, a three-element vacuum tube (or ''thermionic'' valve). Dubbed the de Forest Audion, it was a device that could detect and electronically amplify radio and telephone signals. In 1909 the American company Bell American Telephone & Telegraph (AT&T) bought de Forest's patent and improved the tube so that it could be used to amplify signals in long-distance telephony. A practical problem was that the vacuum tubes were often unreliable, slow, used too much power, and produced too much heat. For years, researchers in Western countries tried to make a solid-state amplifier—what became the transistor—in an attempt to enable the creation of smaller, faster, less power-hungry electronics.

Early experiments in transistor technology were based on the analogy between the semiconductor and the vacuum tube: the ability to both amplify and effectively switch an electrical signal on or off (rectification). By 1940, Russell Ohl at Bell Telephone Laboratories, among others, had found that impure silicon had both positive (*p*-type material with holes) and negative (*n*-type) regions. When a junction is created between *n*-type material and *p*-type material, electrons on the *n*-type side are attracted across the junction to fill holes in the other layer. In this way, the *n*-type semiconductor becomes positively charged and the *p*-type becomes negatively charged. Holes move in the opposite direction, thus reinforcing the voltage built up at the junction (Figure 12). The key point is that current flows from one side to the other when a positive voltage is applied to the layers (''forward biased'').

The transistor is also a solid-state device that can amplify electrical current. Transistor stands for transit resistor. Its development went along two lines: basic research in solid-state physics to replace the old vacuum tubes (such as de Forest's Audion), which failed to solve technological problems, and multidisciplinary research activities at several industrial and university research labs. It was in December 1947 that John Bardeen and Walter Brattain working at Bell Telephone Laboratories in New Jersey, in a research team headed by

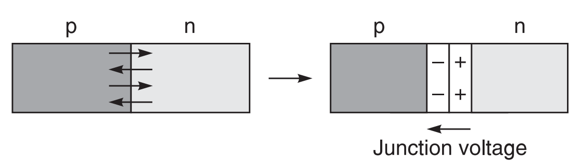


Figure 12. *p–n* junction between two semiconducting materials. Near the junction, electrons on the *n* side diffuse to the *p* material on the left and recombine with holes in that side, and holes diffuse from the *p* side to the *n* material on the right, recombining with electrons. At equilibrium a thin depletion region forms, with an overall negative charge in the *p* material and an overall positive charge in the *n* material. A junction voltage develops across the depletion region.

William Shockley, demonstrated the first transistor, a semiconductor device based on germanium. However, the German scientist Julius E. Lilienfeld from New York had patented the first field-effect transistor in 1926. It was a patent on a "method and apparatus for controlling electric currents." It was unlikely, however, that he ever got it to work. Nevertheless, in the early 1930s solid-state physics, and later semiconductor technology, was a promising research field for a broad range of researchers. Semiconductors were interesting to the radio and telephony industry because of their ability to rectify electrical current (allowing current to flow in one direction and not the other) and they were useful as electronic switches.

In the late 1930s, Bell's director of research Mervin Kelly recognized that a better amplifier was needed for the telephone business. He gave a group of researchers headed by Shockley the freedom to carry out scientific work in the field of solid-state physics. In 1939 Shockley further developed the principle of the field-effect transistor (FET), in which instead of the wire "grid" of de Forest's triode, an electric field controlled the stream of charge carriers between electrodes. This principle started a line of inquiry that led to new experiments. However, Shockley himself went off in other directions and was barely involved in the further experimental research of the Bell group. Meanwhile Robert W. Pohl and Rudolf Hilsch from Gottingen University made a solid-state amplifier in 1938 using salt as the semiconductor. It was a functioning device, but reacted to signals too slowly. Karl Lark-Horovitz and his research team at the physics department of Purdue University, Indiana also became involved in solid-state physics, working on improving the crystal rectifiers that were used as radar detectors in World War II. The team at Purdue worked with both silicon and germanium crystals. Such crystal detectors had no signal gain, but the work on germanium and techniques of growing and doping semiconductor crystals were important to later semiconductor researchers.

The Bell researchers did the most extensive work on crystal rectifiers in the radar program both in the U.S. and the U.K. In 1945 John Bardeen, a theoretical solid-state physicist, joined Shockley's group and a semiconductor subgroup was formed within Bell Laboratories. Shockley filled out his team with a mix of physicists, chemists, and engineers. In this subgroup Walter Brattain was an experimental physicist. Bardeen and Brattain continued the research on Shockley's earlier design sketches for the field-effect transistor. The substitution of the Fleming triode tube (developed by Ambrose Fleming in 1904) by a solid-state device—the transistor—formed the most important outcome of this semiconductor subgroup's research efforts.

In their experiments they placed the electric circuit contacts on two strips of golden foil, since Bardeen suggested that greater amplification could be obtained by placing the two point contacts closer to each other. Bardeen also suggested replacing silicon with high-purity germanium that made better rectifying contacts. Germanium is an *n*-type semiconductor (excess of electrons), and when current flowed in from the gold foil contact, holes were "injected" into the germanium surface. This created a *p–n* junction as described above. In the junction, current started to flow from one side to the other. In the case of their little construction, current flowed towards the second gold contact. The outcome was that a small current changed the nature of the semiconductor so that a larger, separate current started flowing across the germanium and out of the second contact. In other words, a small current was able to alter the flow of a much bigger one, thus effectively amplifying it. The first device was called a point-contact transistor because the wires stood directly in contact with the semiconducting material (Figure 13). Later Shockley developed the junction transistor (also called the sandwich transistor), of which there are two types, called *p–n–p* and *n–p–n* (depending on which material forms the inside layer). The field-effect transistor was not built until the 1960s, but today most transistors are field-effect transistors.

In 1956 Shockley and his colleagues shared the Nobel Prize for their invention of the point-contact transistor. Inner competition broke the Bell Lab team apart, but their invention was of great importance for Bell, of which numerous patents and licenses with amongst others General Electric, IBM, Texas Instruments, Philips, and later Sony Electronics bear witness.

Following Bells announcement in 1948 of the first working transistor, the transistor quickly became popular in industry as Bell licensed their transistor, and the first commercial product with transistors—a hearing aid—was sold by Raytheon in 1952. Military applications, as a replacement for the vacuum tube in communications and computing, swiftly followed. Transistors began to replace fragile vacuum tubes in consumer electronic devices, and the first U.S. transistor radio was sold by Texas Instruments in 1954. Sony Electronics especially was able to mass-produce miniaturized transistor radios from 1957. The

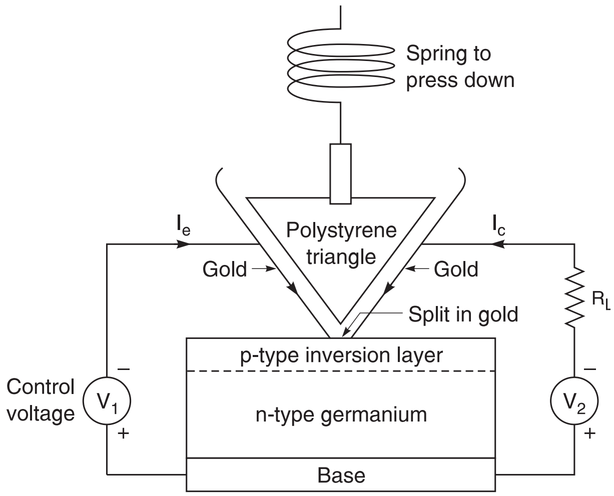


Figure 13. The first point-contact transistor, developed by Walter Brattain and John Bardeen in 1947. A single strip of gold foil over a plastic triangle is split to make two contacts. The triangle lightly touches the surface of a crystal of germanium, which sits on a metal plate attached to a voltage source. Current following from one gold contact to the other can be controlled by current introduced into the germanium. The actual device was about half an inch high. [*Reproduced with permission from Hoddeson, L. Innovation and basic research in the industrial laboratory: the repeater, transistor and Bell Telephone System, in Between Science and Technology, Sarlemijn, A. and Kroes, P. Eds. North Holland, Amsterdam, 1990.*]

transistor became the key to further developments in electronic technology and the consumer electronics industry. Considerable developments were made in the 1950s as a result of open sharing of technology between various industrial and university labs. These developments, like the means of introducing dopants (impurities) to very shallow depths using vapor phase diffusion, the use of silicon dioxide as a diffusion mask, and an all-diffused silicon transistor enclosed in oxide, led to a wide range of transistors.

A practical problem remained, however. Like the elements in the vacuum tubes, the electric components that formed the transistor needed to be soldered together. The more complex the electric circuits became, the more complicated the construction of the transistor. Computer technology in particular needed complex circuits. Because of this problem, practical application of transistors was slowed down. However, in 1958 Jack Clair Kilby of Texas Instruments developed the first integrated circuit or chip using some key achievements from the 1950s transistor research activities. His invention combined a collection of transistors arranged on a single chip of silicon in order to save space. This was the first step to integrated circuits that replaced individual transistors in computers—a refinement that led to the development of the modern microprocessor.

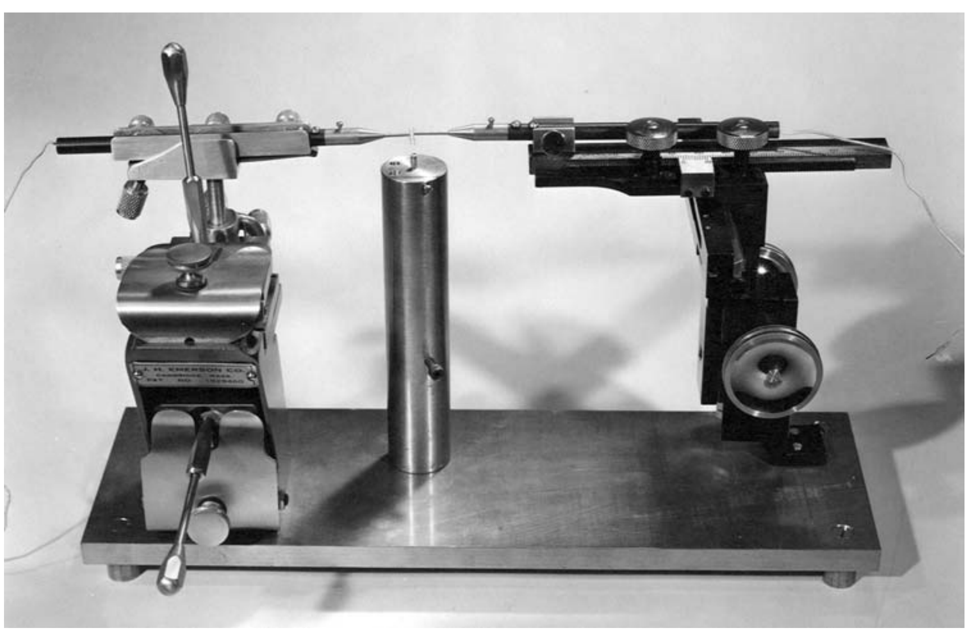


Figure 14. RCA transistor 1948.
[*Courtesy of the David Sarnoff Library.*]

See also **Integrated Circuits; Radio Receivers, Early; Radio Receivers, Valve and Transistor Circuits; Semiconductors; Valves/Vacuum Tubes**

KEES BOERSMA

Further Reading

Bardeen, J. The early days of the transistor. *Proc. Stocker Symp.*, Raju, G.V.S., Ed., 1979, pp. 3–10.
Brattain, W.H. Genesis of the transistor. *Phys. Teacher*, 6, 108–114, 1968.
Braun, E. and Macdonald, S. *Revolution in Miniature: The History and Impact of Semiconductor Electronics*. Cambridge University Press, Cambridge, 1978.
Brinkman, W., Haggan, D. and Troutman, W. A history of the invention of the transistor and where it will lead us. *IEEE J. Solid-State Circuits*, 32, 12, 1997.
Hoddeson, L. Innovation and basic research in the industrial laboratory: the repeater, transistor and Bell Telephone System, in *Between Science and Technology*, Sarlemijn, A.,P., Ed. North Holland, Amsterdam, 1990.
Riordan, M. and Hoddeson, L. *Crystal Fire: The Birth of the Information Age*. W.W. Norton, London, 1997.
Shockley, W. The theory of *p–n* junctions in semiconductors and *p–n* junction transistors. *Bell Systems Tech. J.*, 27, 435–489, 1949.

Useful Websites

Brinkman, W., Haggan, D. and Troutman, W. A history of the invention of the transistor and where it will lead us: www.sscs.org/AdCom/transistorhistory.pdf
http://www.pbs.org/transistor

Transport

During the twentieth century, human beings witnessed a revolution in transportation technology. The way in which people and goods moved around the world changed dramatically. The nineteenth century, of course, also had its transportation revolution, as railroads and steam ships became dominant, and these forms of mass travel did not disappear during the twentieth century. Transport by sea continued to play a key role in the movement of people and cargoes, as ships became bigger and faster, culminating in giant super tankers used to transport oil around the world. Railroads also evolved, shifted to electric and diesel power. By the late twentieth century, high-speed trains whisked passengers to their destinations at speeds of up to 270 kilometers per hour.

The internal combustion engine led to widespread use of automobiles, buses, and trucks. Trucks, for example, would provide fierce competition for the railroads in the movement of goods across land. The development of aircraft also transformed travel around the globe. The use of airplanes to carry passengers and cargoes greatly reduced travel time. Journeys across the oceans were now measured in hours rather than in days and weeks.

Water Transport

Transportation over water continued to play an important role in the twentieth century. Intense competition among shipping lines had begun in the mid-nineteenth century as immigration from Europe to the Americas grew. Shipping lines sought to build bigger and more luxurious ocean liners in hopes of attracting first- and second-class passengers. By the early twentieth century, the Hamburg–America Line was the largest shipping company, covering 75 routes with 500 vessels. Rivalries among these shipping lines meant that the companies continually put bigger and faster ships into service, leading to much turnover in their fleets, as older vessels quickly became outdated. This competition also led to a price war in the years before World War I. Lower prices led to a notable increase in emigrant traffic in this period.

At the turn of the century, the U.S. sought to become more involved in the transatlantic shipping business. In 1902, J.P. Morgan formed the International Mercantile Marine Company. Morgan either bought or made cooperative agreements with numerous existing shipping lines. His biggest competitor was the Cunard Steamship Company, a British line. In order to combat Morgan, the Cunard line received subsidies and loans from the British government. It was then able to construct two 35,000-ton liners and in 1903 inaugurated direct service from Italy to New York. The Morgan–Cunard rivalry led to a rate war in 1904, which eventually led to the end of Morgan's venture in 1926.

Despite technological advances, sailing ships did not disappear immediately. However, by the early twentieth century, they had mostly been pushed to more marginal trade. There remained some important transoceanic routes for sailing ships, such as guano from South America, grain from Australia and Argentina, and timber from Africa and Asia. These sailing ships continued to operate where large cargo ships could fill their holds and find favorable winds. However, there were only a few holdovers that lasted until after World War II.

Most cargo was carried by tramp steamers, a business dominated by the British. These tramp steamers could carry cargoes of up to 10,000 tons while traveling at 13 knots. They consumed 70 tons

of coal per day. They were some 140 meters long and 16 feet wide and were manned by a crew of 75.

Shipping in the early twentieth century was aided by the construction of the Panama Canal. This canal stretches 65 kilometers from shoreline to shoreline across the Isthmus of Panama, connecting the Atlantic and Pacific Oceans. Along with the Suez Canal, the Panama Canal became one of the two most strategic artificial waterways in the world. The canal opened to traffic in 1914 and was controlled by the U.S. until 2000 when control reverted to Panama. The great advantage of the Panama Canal was that ships no longer had to carry their cargoes and passengers around Cape Horn at the southern tip of South America, thus shortening the journey between the east and west coasts of the U.S. by some 8,000 nautical miles (14,800 kilometers).

Another technological development in the movement of goods around the world by sea was the increased use of tankers. The first tankers had been used in the late nineteenth century to transport oil. There was much concern over the use of tankers, particularly due to the threat of explosions. For this reason, tankers were not allowed through the Suez Canal until 1902. As petroleum became an increasingly important product in the early twentieth century, tanker freights increased dramatically. At first, most tankers were owned by the oil companies themselves. Later, private companies became involved, leasing their tankers to the large oil firms.

World War II saw important developments in the movement of people and goods across the seas. In particular, shipbuilding in the U.S. was a key factor in the Allied victory. U.S. shipyards produced an enormous fleet that kept the Allies supplied with constant convoys of materials and troops across both the Atlantic and Pacific Oceans. The most important of the ships produced during the war were the Liberty ships, of which 2650 were built between 1941 and 1944. Among the advances seen in the construction of the Liberty ships was that large sections were produced and then welded together. They had steel decks, and were powered by reciprocating steam engines fed by oil-fired boilers.

The postwar years saw an important evolution of shipping technology, as larger tankers and bulk carriers made their appearance. These developments were linked to a significant increase in world trade following the war. Ships became much faster, reaching speeds of 20 knots. There was also much concern over appearance as esthetically pleasing ships were built. Shipbuilders also tried to streamline the loading and unloading of cargo, introducing for example the loading pallet in 1958.

Perhaps the key development in the movement of goods by sea in the postwar era was the fact that oil became the key to the world economy. By the early 1950s, about 40 percent of sea trade was in petroleum products and the total amount of crude oil and gasoline shipped by tanker nearly tripled in the decade. There was already a plentiful supply of 16,000 deadweight ton (dwt) tankers built during the war. However, as oil consumption continued to grow, there was increased demand for higher capacity tankers. Soon, shipbuilders were constructing 50,000 dwt supertankers, which greatly reduced the cost of transport. These supertankers led to many other technological changes, including the need for larger building berths in shipyards, larger dry docks in repair yards, the construction of new terminal berths and storage facilities by oil companies. Changes in shipbuilding were often led by the Japanese, who pioneered the mass production of tankers in their shipyards, using large cranes to lift huge subassemblies and taking the lead in the use of computers in shipbuilding. At first the 50,000- to 60,000-dwt tankers were seen as the limit in size, as larger tankers could not pass through the Suez Canal. However by 1970, there were tankers of 200,000 to 300,000 dwt.

Railroads

By the early-twentieth century, railroads had reached maturity. At the time of World War I, there were about 1.6 million kilometers of railroad track around the world, with some 25 percent located in the U.S. During the first half of the twentieth century, new railroad construction slowed in most of the developed world. However, in other parts of the world, the building of new railways continued. Such was the case in Canada, China, Russia, and parts of Africa. A prominent example of this new construction was the completion of the Trans-Siberian Railroad in Russia. Work had begun on the railroad in 1891. In 1904, all of the sections between Moscow and Vladivostok had been completed. However, Russia had secured permission from China to build part of the line across Manchuria. After the Russo–Japanese War in 1904–1905, Russia feared Japan might take over Manchuria, and proceeded to build an alternative route across Russian territory known as the Amur Railway. This section was completed in 1916.

Another impressive railroad engineering feat of the early twentieth century was the completion of

the Simplon Tunnel in the Alps, which despite many serious problems was opened to railroad traffic in 1906. One of world's longest railway tunnels at 20 kilometers long, it connected Iselle, Italy to Brig, Switzerland. The Simplon Pass had been an important European trade route for centuries. In the 1890s, a German engineering firm undertook construction of the Simplon tunnel using many new tunneling techniques. Other technological developments to aid railroad transportation in the twentieth century included electric signaling systems and the widespread adoption of diesel traction. By the late 1930s, the use of diesel power had occurred in many places as it proved more reliable and efficient than steam power. In most railroad networks, diesel locomotives are more cost effective than electric ones, which is generally only profitable in very high-traffic areas such as metropolitan commuter systems.

Electrification became more widespread in the second half of the twentieth century. The French pioneered electrification of locomotives with a direct supply of high-voltage alternating current at the industrial frequency. These French developments in turn led to subsequent large-scale electrification programs in countries such as China, Japan, South Korea, the Soviet Union, and India. Such conversions to electric traction were also encouraged by the high price of oil during the 1970s.

As was the case in the early twentieth century, some countries expanded their railroad networks in the latter part of the century. This was particularly true in countries that were undertaking major industrialization projects and where railroads were still the principal form of moving people and goods. Examples include China, the Soviet Union, and India. These countries built new rail lines in order to increase their capacity to carry raw materials to industrial areas of their countries. For example, China doubled the length of its rail network between 1950 and 1990, including numerous routes of up to 800 kilometers that connected coal fields in the west to the ports on the east coast. During the same period, the Soviet Union built 32,000 kilometers of new rail lines. This new construction included the more than 3000-kilometer-long Baikal–Amur Trans-Siberian Railroad that was started in the late 1970s. During the 1990s, India also undertook major railroad construction.

In other parts of the world, there was less new construction. Here, there was an emphasis on improving rail communications, especially in light of growing competition from highway construction and air travel. Railroad companies sought to improve passenger amenities, develop larger and more specialized freight cars, design more sophisticated signaling and traffic control, and improve motive power.

Two major railway tunnels built in the second half of the twentieth century were Seikan Tunnel in Japan and the Channel Tunnel connecting England and France. The Saikan Tunnel connects the islands of Honshu and Hokkaido and is the world's longest tunnel at 53.8 kilometers. The Japanese National Railways sponsored construction of the tunnel between 1964 and 1988. The tunnel has quickly become of limited use, as air travel between the islands is faster and almost as cheap. Nearly as long at 31 miles, the Channel Tunnel, popularly called the "Chunnel," provides freight and passenger railroad service on trains that can travel up to 160 kilometers per hour. Digging began in 1987 and the tunnel opened in 1994.

The most significant development in rail transportation was an emphasis on high-speed trains, and Japan pioneered this development. In 1957, the Japanese government concluded that the old Tokyo–Osaka line could not be upgraded to meet the needs of the heavily populated and industrialized Tokaido coastal belt between the two cities. In 1959, work began on a 500-kilometer-long line for electric passenger trains. The first *Shinkansen* (New Trunk Line) opened in 1964. Initially able to travel at speeds of 210 kilometers per hour, the high-speed line was a commercial success. From the 1970s through the 1990s, Japan extended its network of high-speed trains, some of which reached speeds of 274 kilometers per hour. In Europe, France took the lead in developing a system of high-speed trains. The first *Train à Grande Vitesse* (TGV) between Paris and Lyon went into service in 1981, the result of some twenty years of research. France built and planned additional lines with an ultimate goal of connecting all major cities with Paris as well as connecting with the high-speed systems in neighboring countries. Other European nations, including Italy, Germany, and Spain have also developed high-speed trains.

Trucks

Railroads faced competition from commercial trucking throughout the twentieth century. Gottlieb Daimler built the first motor truck in Germany in 1896. In the U.S., truck transportation began at the turn of the century. The Winston Company of Cleveland, Ohio began building

trucks in 1898, making it one of the first truck manufacturers in the U.S. In 1904, there were a mere 700 trucks in use. This figure had increased to more than 150,000 by 1915. The first transcontinental coast-to-coast truck trip occurred in 1911, completed in 66 days. Early trucks were used mainly for making local deliveries and limited commerce between cities. Their usefulness was hindered by poor roads and the dominance of railroads in shipping goods over long distances.

However, before trucks could compete effectively with railroads, the country needed a system of well-paved roads. In 1916, the U.S. government passed the Federal Aid Road Act, which emphasized the construction of high-quality hard-surface pavements and highways for motor transportation, which helped intercity commerce utilizing trucks. As with other technologies, World War I played a key role in the development of motor trucks. In 1914, annual U.S. truck production was 24,900. By 1917, the country was producing 128,000 trucks. Many of these trucks went to Europe for the war effort to move soldiers and supplies and serve as ambulances. Because railroad arteries were often clogged during the war, large truck caravans were organized to drive the new trucks to East Coast ports from assembly plants in the Midwest. The trucks were also loaded with goods, demonstrating the possibilities for long-distance trucking and also calling further attention to the need for system of highways.

Air Transport

The use of aircraft would drastically alter the way in which people and goods traveled around the world during the twentieth century. Experiments with flight were not new in the twentieth century. The French had pioneered the use of balloons in the late-eighteenth century. In the late nineteenth century, the German Otto Lilenthal successfully flew gliders. Others had begun experiments with lighter-than-air dirigibles. The German Count Ferdinand von Zeppelin carried out the most important work, including the creation of a rigid but light frame that allowed for easier steering. Following Zeppelin's advancements, in 1910 the German company Deutsche Luftschiffahuts AG (Delag) was organized and took the lead in transporting passengers in its airships. Between 1910 and 1914, Delag carried more than 34,000 passengers. Such service continued after World War I, when the *Graf Zeppelin* flew more than 1.6 million kilometers in commercial service. However, commercial lighter-than-air service effectively came to end when the zeppelin *Hindenburg* exploded in 1937.

Inspired by Lilenthal, the young mechanics Wilbur and Orville Wright in the U.S. built the first aircraft that was heavier than air and able to maneuver in flight. In December 1903 the Wright brothers made the world's first successful flight, success based on the fact that it was a powered, sustained, and controlled flight that carried a human being.

Following the success of the Wright brothers, human beings would be fascinated with aviation, inspiring new designs and expanding the use of aircraft. The first commercial service was between Tampa and St. Petersburg, Florida, but expansion was further hastened by World War I and stories of World War I flying aces who captured the attention of the public. While commercial developments were interrupted by the war, there were many technical developments made during the conflict. During the 1920s, aviation was mainly a sport and form of entertainment. Amelia Earhart and Charles Lindburgh were among the most well-known pilots of this heroic age of flight.

At the same time, the first small commercial airlines began to operate, often to carry mail. The first were formed in 1919 in Germany, France, and Holland. These would be followed by many more during the 1920s and 1930s. Most of these early commercial airlines did not last long, either failing or merging with other airlines. In 1927, John Northrup and the Lockheed Aircraft Company developed the Vega aircraft, which served as model for modern commercial aircraft. The Vega used either 220- or 425-horsepower engines, carried a pilot and six passengers, could travel at speeds of up to 215 kilometers per hour, and had a range of between 800 and 1450 kilometers. Starting in the 1930s, airplanes were used increasingly for the transportation of people and goods. Aviation technology advanced, as monoplanes with all-metal fuselages and retractable undercarriages became widespread. In particular, three major airlines began to establish worldwide routes during the 1930s: Pan American, Imperial, and KLM.

These airlines often used seaplanes. Water provided potentially long runways at a time when most airport runways were only about 300 meters feet long. In turn, this meant that planes could be larger, use multiple engines, and have larger fuel tanks, allowing them to carry more passengers over longer distances. The 1930s also saw the introduction of first true modern commercial aircraft. In 1933, Boeing introduced its Boeing 247, to be followed by the Douglas Company's DC-2 and

DC-3. The DC-3 proved to be the first commercial aircraft to be profitable. It could fly at above 1500 meters, had a stressed aluminum sheathing that gave it considerable strength, and had a retractable landing gear. DC-3s carried most commercial traffic in the U.S. at the time, showing that commercial air travel could be safe, reliable and profitable.

The next technological advancement in the movement of people and goods through the air was the four-engine plane. Work on such aircraft had begun as early as 1913 and demand increased during the 1930s. Airlines sought to fly longer routes, such as across the Pacific Ocean, and only more powerful four-engine planes could lift enough fuel to fly such distances. Four-engine planes could also fly at higher altitudes, thus avoiding the "weather," making trips faster and more comfortable for passengers. They could also fly over mountains rather around them. Another technological advance that had to accompany the four-engine planes was the pressurized cabin. An early example of this new generation of commercial aircraft was the Boeing Stratoliner, introduced in 1940.

War once again played a role in developing aviation technology. During World War II, airplanes became bigger and faster. The war was also important as the first aircraft with jet engines were introduced. After the war, there was tremendous growth of commercial air travel, led by the Stratoliner, the DC-4, and the Lockheed Constellation, all aircraft that were faster and could travel greater distances than earlier planes. By the late 1950s, more people crossed the Atlantic Ocean by plane than by ship. In the 1950s, there was some use of turboprop planes, mostly in Europe. However, the turboprop planes would soon be surpassed by the widespread use of jet engines introduced during the war. By the late 1940s, jet engines were standard in most military aircraft and by the 1950s, they were used increasingly in commercial aviation. Due to the higher speeds and lower operating costs of jet aircraft, there was a major expansion of the commercial airline industry in the second half of the twentieth century. The potential of this technology could be seen in the 1970 introduction of the Boeing 747 "jumbo jet," capable of carrying up to 500 passengers.

Once jets became more widespread, the next technological advance was to produce supersonic aircraft. If successfully developed, airplanes traveling faster than the speed of sound would greatly reduce flight times around the world and revolutionize communications. Starting in 1962, France and Great Britain agreed to jointly develop a supersonic transport (SST0), known as the Concorde. The first passenger plane to break the sound barrier, however, was the Soviet Tu-144. Due to design problems, however, the Soviet craft was only in service briefly during the 1970s. The Concorde made its first flight in 1969 and entered service in 1976. It reduced the flight time between London and New York to about 3 hours. However, supersonic travel did not become profitable and also met resistance from environmental groups. In 2000, a Concorde crash outside Paris led many to reconsider the safety and value of the supersonic fleet. The Concordes were withdrawn from service because of high fuel and maintenance costs in late 2003.

See also **Aircraft Design; Bulk Carriers and Tankers; Rail; Urban Transportation**

RONALD YOUNG

Further Reading

Allen, G.F. *Railways of the Twentieth Century*. W.W. Norton, New York, 1983.

Allen, G.F. *Railways: Past, Present, and Future*. Morrow, New York, 1982.

Bilstein, R. *Flight in America: From the Wrights to the Astronauts*, revised edn. Johns Hopkins University Press, Baltimore, 1994.

Heppenheimer, T.A. *A Brief History of Flight: From Balloons to Mach 3 and Beyond*. Wiley, New York, 2001.

Heppenheimer, T.A. *Turbulent Skies: The History of Commercial Aviation*. Wiley, New York, 1995.

McCullough, D. *Path Between the Seas: The Creation of the Panama Canal, 1970–1914*. Simon & Schuster, New York, 1977.

Nersian, R.L. *Ships and Shipping: A Comprehensive Guide*. Penwell Books, Tulsa, 1981.

Rinman, T. and Brodefors, R. *The Commercial History of Shipping*. Translated by Hogg, P. Rinman and Linden AB, Goteborg, 1983.

Transport, Foodstuffs

Twentieth century foodstuffs were transported by land on vehicles and trains, by air on cargo planes, and by water on ships or barges. Based on innovations used in previous centuries, engineers developed agricultural technology such as refrigerated containers to ship perishable goods to distant markets. Technological advancements enabled food transportation to occur between countries and continents. International agreements outlined acceptable transportation modes and methods for shipping perishables. Such long-distance food transportation allowed people in different regions

of the world to gain access to foodstuffs previously unavailable and incorporate new products they liked into their diets.

Land and ocean transportation methods are the most economical methods for food shippers. Because of expenses, aircraft have been used to transport food primarily when profits from sale of goods are expected to exceed shipment costs or urgency is involved such as the humanitarian Berlin Airlift. Since the mid-twentieth century, technology enabled foods to be safely preserved and packaged for transportation to serve as meals for passengers on aircraft and boats. Foods undergo such processes as being heated, cooled, or hydrated in order to be palatable. As a result, airplane passengers can dine on fresh fruit while flying 10 kilometers high above frozen lands in the winter. People on cruise ships can feast on fresh nonseafood meat and dairy products and edible vegetables when they are hundreds of kilometers from shore.

Preservation of the quality of foodstuffs en route from field to market is crucial to maintain flavor, texture, and appearance. Temperature and humidity regulation maintained by air circulation within shipping devices is essential for foodstuffs transportation. Most types of vegetables, fruits, meat, and dairy and poultry products require chilling or refrigeration. Humidity is maintained at 85 to 95 percent to prevent foods from dehydrating. Foodstuffs are usually loaded in large metal intermodal containers which are standardized 20- to 40-feet (6- to 12-meter) boxes that can be pulled as trailers by trucks on highways, carried by ships, or moved by railways as a trailer-on-flatcar (TOFC). In the early twentieth century, Mary Engle Pennington designed refrigerated rail cars that were adopted internationally. Container innovations included building a double hull to insulate products and minimize leaks and contamination.

Refrigerated trailers dominate road food transportation methods. This transportation mode minimizes food vulnerability to shipment damage from being harvested to placement on grocery shelves. Refrigerated transport enables fresh produce from milder climates to be shipped out-of-season to colder locations. Refrigeration is achieved by mechanical or cryogenic refrigeration or by packing or covering foods in ice. Ventilation keeps produce cool by absorbing heat created by food respiration and transferred through the walls and floor from the external air beneath and around the shipping trailer.

Engineers designed trailers to have one of two types of air circulation, bottom forced-air delivery or overhead top-air delivery. Most refrigerated trailers have overhead top-air delivery. A blower pushes cold air through ceiling ducts towards the rear of the trailer. The air moves horizontally through channels between packed items then downward, passing underneath the load before returning to the blower. Walls are grooved and floors have T rails or pallets so air flow is not obstructed. Vents remove heat from the circulating air.

Bottom forced-air delivery is most frequently used on ship containers although some over-the-road trucks have this type of air circulation. Air moves from a container's bottom, passing under and up through loads. Shipment boxes have vent holes in their tops and bottoms and are packed tightly so that the holes match up to facilitate air passage. Transportation vehicles are heated if the fresh foods being shipped risk frost damage when they pass through freezing weather conditions.

Tankers carry milk, oils, and other liquid foodstuffs. Live animals shipped for food consumption require specific transportation needs for humane and public health reasons. Special tanker trucks are designed to haul live fish between fisheries and markets. Other vehicles ventilated with air holes and equipped with ramps are built to transport cattle and poultry from farms to slaughterhouses.

Trailer-loading techniques aid ventilation, which keep foods at optimum hauling temperatures. Shippers prepare food for transit by precooling it according to industrial standards. Preferably, heat is removed immediately after perishables are harvested. Various methods are utilized, including application of liquid ice, hydrocooling, vacuum cooling, and refrigerated forced air chilling rooms. Foodstuffs can also be refrigerated by vehicles at loading docks. Prechilling results in foods retaining flavor, freshness, and appearance when they reach market and inhibits decay thus increasing the shelf life.

Food technologists design packaging materials for food transportation. Most produce is shipped in corrugated and fiberboard cardboard boxes that are sometimes coated with wax. Wooden and wire-bound crates are also used in addition to bushel hampers and bins. Mesh plastic, burlap, and paper bags hold produce. Meat is often vacuum packed on plastic trays that are placed in wooden lugs. Foods are occasionally wrapped in plastic liners or packed in ice to withstand damage in transit and limit evaporation.

Loading patterns are essential to achieve desired circulation to maintain desired shipping temperatures. Boxes are stacked to create horizontal and vertical air channels while ensuring that the load is

stabilized to prevent damage from road vibrations and movements. Certain foods such as dairy products are lifted above the floor to prevent them melting due to road heat. Eggs are shipped in cartons placed in fiberboard boxes, which are packed tightly together to prevent breakage. Modular unitized metric (MUM) load standardizes loading unitized produce packaged together on pallets and reinforced with straps. Units are quicker to load and unload than individual cartons, and handling damage is minimized.

Approximately 5 percent of fresh foodstuffs are ruined during distribution. Some fruits and vegetables produce ethylene gas that can damage other foods. Genetic engineers have designed produce that is more tolerant of transportation stresses. Digital information and electronic alarms alert shippers to the temperature and humidity of shipments through modems and satellite links, and remote controls provide access to alter settings. Inside shipping containers, a controlled atmosphere is maintained. Engineers determined that removing oxygen prevents foods from decaying while in transit. Carbon dioxide or nitrogen gases are injected into the container to supplement temperature and humidity control. After produce is loaded, shippers seal the container's contents with plastic film. Air valves are shut, and gases are released. Carbon dioxide is used to inhibit mold in berries. Nitrogen protects green vegetables and fruits. Small loads are sealed in plastic bags with protective gases.

Ocean transportation of food is aided by international deep-water cargo ports connected to railroad yards. In 1966, SeaLand was the first company to utilize shipping containers to trade between Europe and the U.S. Transported in stacks on ships, containers are transferred to trains or trucks or stored in terminals. Globally, containers became accepted during the early 1970s, annually carrying millions of tons of cargo by the 1980s. Until refrigeration became economically feasible in the mid-twentieth century, fish were too perishable to ship beyond local markets. On refrigerated factory ships, filleting and processing occur at sea soon after fish are caught. Seawater mixed with tetracycline antibiotics are used to refrigerate fish and prevent bacterial growth. This reduces spoilage and lost profits in addition to expanding market possibilities.

See also **Food Preservation: Cooling and Freezing; Food Preservation: Freeze Drying, Irradiation, and Vacuum Packing**

ELIZABETH D. SCHAFER

Further Reading

Ashby, B., Hunt, R., Hinsch, T., Risse, L.A., Kindya, W.G., Craig Jr., W.L. and Turczyn, M.T. *Protecting Perishable Foods During Transport by Truck*. U.S. Department of Agriculture, Agricultural Marketing Service, Transportation and Marketing Division, Washington D.C., 1995.

Coyle, W. and Ballenger, N. *Technological Changes in the Transportation Sector: Effects on U.S. Food and Agricultural Trade*. U.S. Department of Agriculture, Economic Research Service, Washington D.C., 2000.

Friedland, K. and Koninckx, C., Eds. *Maritime Food Transport*. Böhlau Verlag, Köln, 1994.

Harris, T.C. *Inland Challenges Facing Deep Sea Carriers in Europe*. Intermodal, Rotterdam, 1998.

Heap, R., Kierstan, M. and Ford, G., Eds. *Food Transportation*. Blackie Academic & Professional, London, 1998.

Kochersperger, R.H. *Food Warehousing and Transportation*. Lebhar-Friedman Books, New York, 1978.

Kochersperger, R.H. *Traffic and Transportation: Servicing the Grocery Industry*. National American Wholesale Grocers' Association, Cornell University, Ithaca 1990.

Miles, G.H. *Alternative Food Delivery Systems: An Exploratory Assessment*. National Science Foundation, Washington D.C., 1977.

Walker, M.S. *Chilled Meat Technical Cold Chain Challenges*. Intermodal, Rotterdam, 1998.

Transport, Human Power

Human-powered transport in the form of bicycles and tricycles has played an important—though often unrecognized—role in developments since the nineteenth century in industrial practices, human mobility, and sporting achievement. It has stimulated engineering innovation far beyond its own boundaries. Henry Ford and the Wright Brothers all began their careers as bicycle mechanics, and bicycles continue to be used because of their simplicity of design as a testing ground for new innovations in materials and bonding methods.

The contemporary bicycle can be traced to the German Baron von Drais, who in 1817 invented a running machine or hobby horse which became popular in France and the U.K. The two-wheeled machine was propelled by straddling the frame and pushing one's feet against the ground. Over subsequent decades, a variety of hand-propelled or other models were built by engineers and craftsmen, but the birth of the bicycle industry is generally dated from when the Parisian carriage maker Pierre Michaux (or possibly his associate Pierre Lallement) added pedals to the front wheel of an old hobby horse in the mid-1860s.

Several different lines of innovation continued simultaneously through the 1870s and 1880s.

James Starley of the Coventry Machinists Company, transformed the heavy iron French "boneshaker" into the lighter steel machine now known as the "Penny Farthing." Starley increased the size of the front wheel to cater to the needs of racing cyclists wanting as much leg power as possible, though other potential users found the high-wheeled bicycle either too dangerous or—for ladies—too indecorous. Other innovations improved drive train efficiency, steering and rolling performance, whilst others focused on safety, compensating for the reduced power of a smaller front wheel by introducing various gearing devices, or by turning instead to tricycles. Innovations such as ball bearings and tangent spoking remain a feature of contemporary bicycles, whilst this period also introduced differential gearing, later transferred into automobile technology. The two key breakthroughs came with the Rover safety bicycle of John Kemp Starley (James's nephew), which attached gearing to the rear, rather than to the front, wheel and the pneumatic tire invented by John Boyd Dunlop from Dublin. During the 1890s, bicycle design stabilized around this low-wheeled, rear chain-driven, pneumatic tired configuration—this has remained the dominant bicycle design to the present day.

During the late nineteenth century, the bicycle was significant as a symbol of modernity, freedom, and mobility; it helped transform the urban landscape and it expanded the horizons of the poor, the rural population, and women. In the twentieth century this symbolism became attached to the automobile, and many cycle manufacturers moved into automobile production. Some, such as Starley's Rover Company, eventually abandoned bicycles altogether. Nevertheless, in Europe the bicycle remained a popular means of both mass transportation and mass leisure until after World War II. The remaining manufacturers adopted many of the mass production methods such as interchangeability of parts and vertical integration that had been advanced by the Pope Manufacturing Company and subsequently the Ford Motor Company. U.K. firms such as Raleigh, BSA and Hercules introduced sheet steel stamping, automatic machine tools and conveyor belt production lines during the first few decades of the twentieth century.

By the 1960s, the bicycle had become marginalized as a form of transport in the developed world due to increasing car ownership and car-based urban planning. Sports cycling took over from commuter and leisure uses, marked by a shift from the robust upright utility machines of the 1950s to the racing bike, characterized by drop handlebars, thin wheels and 10-speed derailleur gears, the latter replacing the once-popular three-speed Sturmey–Archer hub gear. From the 1960s, further innovations were introduced to counter lost markets. Alex Moulton's small-wheeled suspension bicycle inspired a rash of new everyday utility models using balloon tires instead of Moulton's more expensive rubber suspension device. The development of children's models such as the Raleigh Chopper and children's BMX (bicycle moto cross) sports bikes, followed this trend. In the mid-1970s, mountain bikes were developed by Californian downhill racing enthusiasts, based on children's bicycle frames from the 1950s and motorcycle and BMX components. Mountain bikes were sturdier than racing bikes, with thick tires and flat handlebars. By the late 1980s they had taken a major part of the bicycle market, accounting for over 60 percent of sales.

From the industry's perspective, mountain bikes are most significant in being central to a shift in world production to the Far East, especially Taiwan and Japan. Except for specialist models, bicycles sold in the developed world were generally, by the 1990s, constructed in Taiwanese factories and fitted with components made by the Japanese firm Shimano, whose highly innovative approach revolutionized much bicycle componentry. Indexed gearing became widespread, with gearshifters using ratchet mechanisms adapted from Shimano's fishing rod products, and integrated with the braking system. Other innovations focused on drivetrain transmission and integrated shoe-pedal systems. Shimano's American competitor SRAM introduced the "Gripshift" gearshifter in the mid-1990s, while both companies—as well as the revived Sturmey-Archer—developed in the late 1990s new hub gear mechanisms to attract the growing utility and leisure markets that arose as a result of concerns about fitness and the environment.

Most successful bicycle innovations have been proved in competition. The first bicycle races took place in the 1860s to test competing designs. In the 1930s, challenges to the hour distance record were made on "recumbent" bicycles, where the rider sits back with their legs pedaling in front of them (Figure 15). The greater speeds possible in this position, especially when partial or full fairings are added to reduce wind resistance—led to recumbents being banned from cycling competitions so as not to invalidate other models. In the 1970s, recumbent enthusiasts established their own sporting bodies, with speeds reaching as high as 128

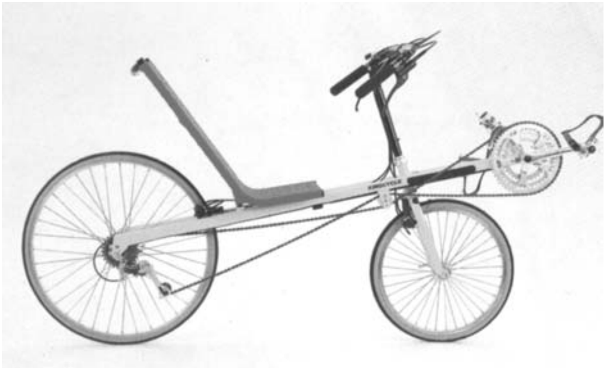


Figure 15. The Kingcycle recumbent, marketed as "the fastest production road bike in the world." [*Source: Neatwork distribution brochure*.]

kilometers per hour in race conditions. Also known as HPVs (human-powered vehicles), there are many competing designs—both bicycles and tricycles are common, and there is little standardization so far for wheel size, componentry, length or height.

The most well known HPV designer—and bicycle innovator more generally—is Mike Burrows, known principally for his very low tricycle design, the Windcheetah. Burrows also designed the Lotus Sport carbon fiber monocoque (single blade) bicycle on which Chris Boardman won the 1992 Olympic pursuit championship for Britain (Figure 16). This machine was designed to minimize aerodynamic drag, matching Boardman's "sports science" approach. A contrasting amateur approach came from the Scot Graeme Obree, who beat several records in the 1990s on a home-made bicycle using washing machine ball bearings, a combination of conventional and found components, and an innovative hunched riding position that was banned by the cycle sport regulators.

Future developments in bicycle design, production and performance are likely to include further improvements in materials and design, although past experience suggests that only in componentry will these filter through to the wider consumer market.

See also **Automobiles; Globalization; Sports Science and Technology; Transport; Urban Transportation**

PAUL ROSEN

Further Reading

Ballantine, R. *Richard's 21st Century Bicycle Book*. Pan Macmillan, London, 2000.

Burrows, M. *Bicycle Design*, Hadland, T., Ed. Open Road, York, 2000.

Dodge, P. *The Bicycle*. Flammarion, Paris, 1996.

Gaffron, P. Cycling special issue. *World Transport Policy and Practice*, 7, 3, 2001.

Hounshell, D. *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States*. Johns Hopkins University Press, Baltimore, 1984.

Lowe, M. *The Bicycle: Vehicle for A Small Planet*. Paper 90, Worldwatch, Washington D.C., 1989.

McGurn, J. *On Your Bicycle: The Illustrated Story of Cycling*, 2nd edn. Open Road, York, 1999.

Norcliffe, G. *The Ride to Modernity: The Bicycle in Canada, 1869–1900*. University of Toronto Press, Toronto, 2001.

Rosen, P. *Framing Production: Technology, Culture and Change in the British Bicycle Industry*. MIT Press, Cambridge, MA, 2002.

Figure 16. Wind tunnel test of the Lotus Sport Superbike ridden by Chris Boardman. [*Source: QED: The Bike, BBC Education, 1993*.]

Useful Websites

Gaffron, P. Cycling special issue. *World Transport Policy and Practice*: http://www.ecoplan.org/wtpp/.

Traveling Wave Tubes

One of the most important devices for the amplification of radio-frequency (RF) signals—which range in frequency from 3 kilohertz to 300 gigahertz—is the traveling wave tube (TWT). When matched with its power supply unit, or electronic power conditioner (EPC), the combination is known as a traveling wave tube amplifier (TWTA). The amplification of RF signals is important in many aspects of science and technology, since the ability to increase the strength of a very low-power input signal is fundamental to all types of long-range communications, radar and electronic warfare.

The traveling wave tube owes its name to its mode of operation: it is designed to cause an RF carrier wave to travel along its length in a carefully predetermined manner. The energy for amplification is derived from a high-powered electron beam, which is made to interact with the RF wave carried on a slow wave structure, usually in the form of a helix (see Figure 17). The helix, made of copper, or of tungsten or molybdenum wire, is supported by three or four ceramic rods to isolate the RF fields on the helix from the metallic walls of the surrounding vacuum envelope. The helix reduces the velocity of the RF wave so that it travels slightly slower than the electron beam. This allows an interaction between the electron beam and the RF wave, whereby electrons are, on average, decelerated by the electric fields of the RF wave and lose energy to it, thus amplifying the signal it carries. The remaining energy of the electron beam is dissipated as heat in a collector. An alternative type of slow wave structure—known as a coupled cavity design—is based on a series of accurately sized and shaped RF cavity sections, usually of copper, which are brazed together and coupled by a slot in the wall of each cavity.

The TWT was invented in 1943 by Rudolf Kompfner, an Austrian refugee working for the British Admiralty. He demonstrated the principle of traveling wave amplification at Birmingham University later that year and published the first results of his work in the November 1946 issue of *Wireless World* magazine. The first practical device was developed at Bell Telephone Labs (BTL) in 1945 by John R. Pierce and L.M. Field and a detailed theory of its operation was published by Pierce in 1947. Subsequent development work was done at BTL and Stanford University in the U.S., and by Standard Telephones and Cables (STL) in the U.K., with a particular eye to potential applications in the communications field. The first TWTs to enter operational service were built by STL for a television relay link between Manchester and Edinburgh, which was operated by the Post

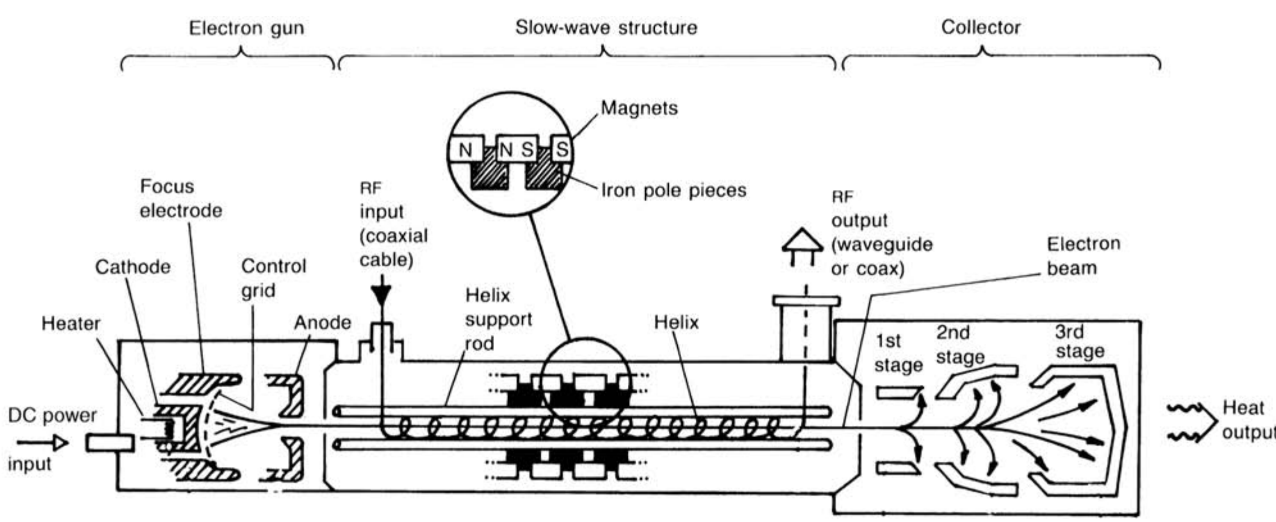


Figure 17. The component parts of a traveling wave tube (TWT). From left to right: electron gun, slow wave structure and collector. Note that the collector of the high-powered space tube depicted is designed to radiate directly into space.
[*Photo: Mark Williams.*]

Office and used by the BBC from 1952. These early tubes produced an output of about 2 watts across a band of frequencies centered on 4-gigahertz and had a gain of about 25 decibels.

Since then, TWTs have been developed to operate at a wide range of radio frequencies and at very high powers, both on earth and in space. In fact, TWTs have been used for space communications (in spacecraft and earth stations alike) since the first privately owned communications satellite, Telstar-1, was launched in 1962. It is historically significant that, without the TWT, the world would not have seen the early Olympic Games or the manned lunar landings "live via satellite."

Today, TWTs are broadband devices that can handle a signal bandwidth up to about 800 megahertz at Ku-band (12 to 18 gigahertz) and even wider bandwidths at higher frequency bands. Ground-based tubes deliver higher output powers than space-based tubes, typically up to 700 watts at Ku-band and 3 kilowatts at C-band (4 to 8 gigahertz) for helix tubes and up to 10 kilowatts for coupled-cavity tubes. Space-based devices, which tend to be of the helix type, can deliver up to about 300 watts at Ku-band. Individual TWTs have been built with power gains of more that 10 million (70 decibels).

In addition to exploiting the TWT for defense communications applications, the military services have developed its potential in the fields of radar and electronic counter-measures (ECM), for which its high gain and broad signal bandwidth are ideally suited. TWTs are also used for civilian and military space-based radars, weather and other remote sensing satellites, and all types of manned spacecraft.

Another type of electron tube, the klystron, was developed at Stanford University in the late 1930s by two brothers, Russell and Sigurd Varian. The klystron power amplifier (KPA) provides a useful alternative to the TWTA in radar systems and satellite earth stations. However, its narrower operational bandwidth and the fact that it cannot be easily retuned has made it unattractive for space applications.

Space-based TWTs are designed to be particularly reliable, since they cannot be repaired once launched. For example, the TWTs on the Voyager-1 spacecraft—at over 12 billion kilometers from the earth, the most distant example of twentieth century technology—were still working 23 years after its launch. Moreover, the ruggedness of the TWT was proved in the late 1970s when a satellite fell 9 kilometers into the Atlantic Ocean following a launch vehicle explosion: the recovered tubes were found not only to work, but to meet their original performance specification.

An important alternative to the TWTA, particularly for communications satellites, is the solid-state power amplifier (SSPA), an RF amplifier which uses semiconductor components—typically gallium arsenide field-effect transistors (GaAsFETs). Although the performance and reliability of SSPAs has been improved since their introduction in the mid-1980s, they are limited to relatively low-power applications at a given frequency since the solid-state medium is a much poorer conductor of heat than the materials used in a TWT. Typical late-1990s GaAsFETs could individually provide up to about 45 watts at C-band and 15 watts at Ku-band, while using power combination techniques, total output powers of 500 watts and 100 watts, at C and Ku-band respectively, were available for certain applications.

The fact that all three types of high-power RF amplifier—TWTA, KPA and SSPA—are still in use, and will remain so for the foreseeable future, indicates that there is often more than one engineering solution to a technological requirement.

See also **Radar, High Frequency and High Power; Radio-Frequency Electronics; Radionavigation; Satellites, Communications; Vacuum Tubes/Valves**

MARK WILLIAMSON

Further Reading

Elbert, B.R. *Introduction to Satellite Communication*. Artech House, Boston, 1999.

Elbert, B.R. *The Satellite Communication Ground Segment and Earth Station Handbook*. Artech House, Boston, 2001.

Evans, B.G. *Satellite Communication Systems*. Institution of Electrical Engineers, London, 1999.

Gittins, J.F. *Power Travelling Wave Tubes*. English Universities Press,. London, 1965.

Hansen, J.W. *Hughes TWT/TWTA Handbook*. Hughes Aircraft Company Electron Dynamics Division, Torrance, CA, 1993.

Kompfner, R. *Wireless World*, 52, 369–372, 1946.

Pierce, J. *IRE. Proc*., 35, 111–123, 1947.

Pierce, J. *Traveling Wave Tubes*. Van Nostrand, 1950.

Williamson, M. *The Communications Satellite*. Adam Hilger/Institute of Physics Publishing, Bristol, 1990.

Tunnels and Tunneling

The history of tunnel construction goes back to the ancient civilizations of the Incas, Babylonians, Persians, and Egyptians, and therefore considerable experience in the construction of tunnels had already been gained worldwide by the beginning of

the twentieth century. Tunnels were constructed to allow transportation through barriers (mountains, underground or underwater). In a country such as Switzerland or Canada, of which substantial parts are mountainous, tunnels were crucial for the development of a transportation infrastructure, and by the end of the nineteenth century the number of railway tunnels had greatly increased.

The optional methods for constructing tunnels increased in the twentieth century. The development of new methods and the improvement of existing ones were stimulated by the rapid increase of car traffic and the need for roads, for which new tunnels were needed. The choice for a particular way in a certain situation depends on the sort of material through which the tunnel is to be constructed. The most important difference is between hard rock and soft material. Besides that, the length and diameter of the tunnel has an influence on this choice. For allowing a sophisticated choice, geologic investigations into the behavior of the ground mass and the ground water are needed in an early stage of the tunnel project.

By the end of the twentieth century the following methods for tunnel construction could be distinguished:

1. Advance the heading in full face without lining. In this case, holes for explosives are drilled, and after blasting the debris is removed. This method is suitable when constructing tunnels through mountains and when the rock is stable enough to be self-supporting.
2. Classical methods. In this process the tunnel cross-section is gradually extended (contrary to the first-mentioned method) and temporary supports are constructed to prevent bursting of the tunnel wall.
3. Trenching or "cut and cover" methods. In this method a trench is excavated and, depending on the behavior of the groundwater, shields are constructed to prevent collapsing of the trench (Figure 18). In case of a stable ground, stiffening ribs may suffice. After constructing the tunnel, the trench has to be covered again.
4. Tunnel boring with a slurry shield. In this case a tunnel-boring machine (TBM; Figures 19 and 20) is used to excavate the tunnel and the walls are temporarily covered with bentonite in order to prevent collapsing, after which a series of tunnel segments is positioned.

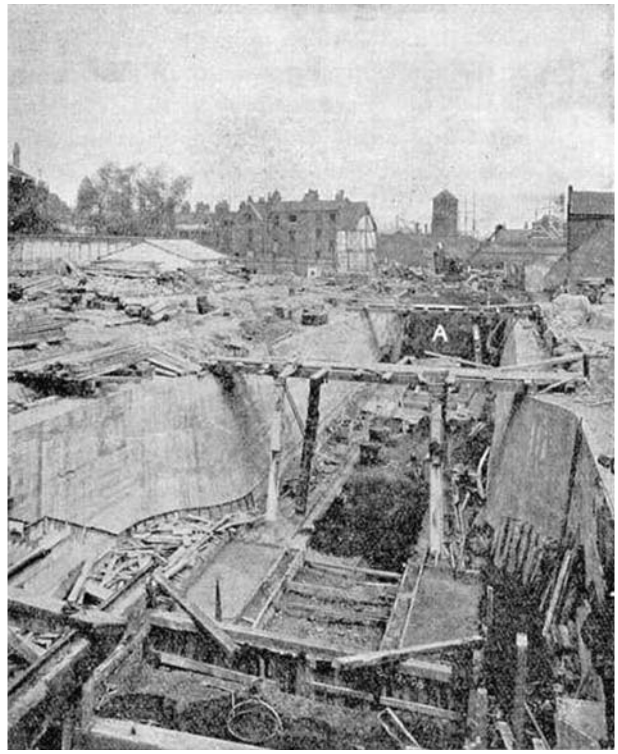


Figure 18. Trenching method as used in the Rotherhithe tunnel. This tunnel was built under the river Thames in England from 1904 to 1908. Two trenches of about 4 meters wide were dug, and side walls built in them. The earth was then removed (letter A in photo indicates where it was dumped).

5. Immersed tunnels. This method can be applied when a tunnel is to be constructed underwater. A prefabricated tunnel segment is positioned at surface level and then sunk, after which the segments are joined.

Of these methods the first had already been fully developed by the beginning of the twentieth century. The second method was improved in the twentieth century by reducing the cost for temporary supports (e.g., by using sprayed concrete). The third method had also been developed in the second half of the nineteenth century, mainly for constructing underground railways in cities (starting with London). The last two methods are twentieth century accomplishments, although there are some roots in the nineteenth century. One of the first tunnels constructed by boring with a slurry shield was the Holland tunnel under the Hudson River in New York, which was opened in 1927. In 1825 the engineer M.I. Brunel had tried to construct a tunnel under the Thames using this method, but this effort had been so problematic that the method remained

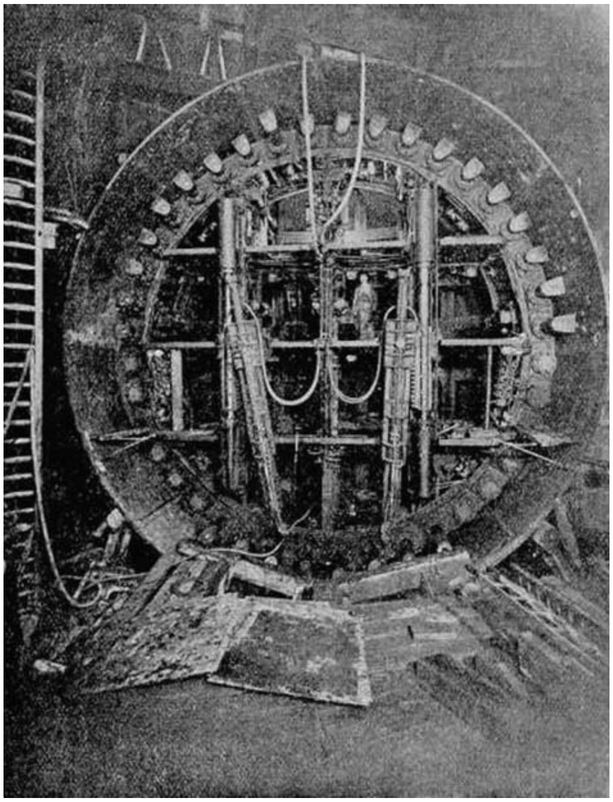


Figure 19. Tunnel boring for the Rotherhithe tunnel. Part of this tunnel was bored with a rotary excavator, a hollow cylinder of 9.5 meters in diameter and 5.5 meters long. The front part (visible in the figure) was built up of cast-steel segments. Also shown are the two erectors for positioning the lining plates.

unpopular until the boring machines were improved. In 1970 for the first time more tunnels were constructed by boring than by blasting. In 1910 the first immersed tunnel was constructed: the Michigan Central Railroad tunnel under the Detroit River. The first immersed tunnel in Europe was built in the city of Rotterdam, the Netherlands (the Maas tunnel completed in 1942). Sometimes different methods were combined, such as in the case of the Detroit–Windsor tunnel that was constructed from 1928 to 1930 by using a combination of trenching and immersing.

The choice for a certain method is not only influenced by geologic considerations, but also by economic and social factors. Of course the methods mentioned above vary in cost and efficiency. Tunnel boring machines in general cost less time than other methods. In hard material, boring with a TBM can yield a progress of up to 15 meters per 8-hour shift, while methods using explosives usually do not yield a progress higher than 7 meters per shift. On the other hand the boring method requires an expensive TBM, which can undo the cost economic effect of the high speed of progress. The influence of the construction process at surface level also has an impact on the choice. Trenching in some cases may be cheaper than boring, but totally upsets the surface level situation, while boring can be done in such a way that it is hardly noticed at surface level (which of course is quite a social advantage). One of the most important social factors for the construction and use of tunnels is safety. Accidents such as a tunnel breakdown or a fire can totally isolate people from the outside world. Also, the air in the tunnel can be easily polluted unless special measures are taken to ensure constant refurbishment of fresh air. Lighting is another important safety issue. In the twentieth century the awareness of these nontechnical aspects grew when the number of tunnels increased and with that the number of accidents.

One of the major achievements in twentieth-century tunneling is the Simplon tunnel (Switzerland), the construction of which commenced in 1898 and was finished in 1905.

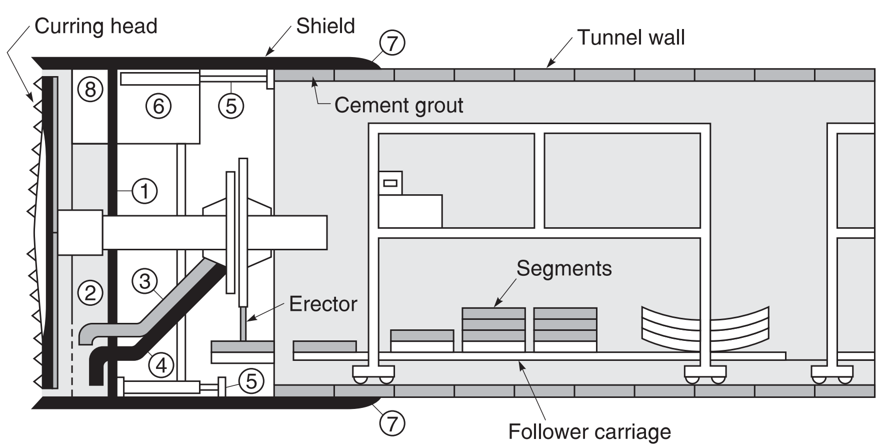


Figure 20. The tunnel-boring machine (TBM) with slurry shield method. (1) Pressure bulk head; (2) baffle; (3) bentonite supplies; (4) bentonite disposal; (5) press screw; (6) air lock; (7) tail seal; and (8) air cushion.

Hydraulic drills were used for removing the mountain rock. A serious delay was caused by a subterranean river that workers encountered in 1901. It was a critical time for the project because all the available materials for constructing girders seemed inefficient to resist the large forces. After only 45 meters, circumstances improved, but it had cost nearly £1000 per yard to overcome this distance. Hot springs that caused high temperatures in the tunnel were another big challenge for the engineers. On 24 February 24 1905 the Italian and the Swiss parts of the tunnel were connected with a heavy charge explosion. The excavating direction had been extremely accurate for that time (up to just centimeters). The tunnel was 19.2 kilometers long and was to be the longest for many years, and it also deserves mentioning that in the construction process only 39 lives were lost.

Some examples of other major tunnel projects in the twentieth century are the Channel tunnel that connects England and France, and the so-called Big Dig project in the city of Boston, Massachusetts, begun in the 1980s and scheduled for completion in 2004. The history of the Channel tunnel offers a nice illustration of political aspects of tunnels. The plans for this 50-kilometer tunnel for many decades were frustrated by the fear of the political effects of a connection between the two countries. The Big Dig project nicely illustrates the enormous impact tunnel constructions can have on life at surface level. By the end of the twentieth century the world's longest completed road tunnel was the Laerdal tunnel in Norway, which was opened on 27 November 2000 (24.5 kilometers in length). The longest railway tunnel was built in Japan (the Sei-kan tunnel, opened in 1988, nearly 54 kilometers length).

See also **Construction Equipment; Urban Transportation**

MARC J. DE VRIES

Further Reading

Beaver, P. *A History of Tunnels*. Peter Davies, London, 1972.

Bickel, J.O., Kuessel, T.R. and King, E.H., Eds. *Tunnel Engineering Handbook*, 2nd edn. Chapman & Hall, New York, 1996.

Harris, F. Ground Engineering Equipment and Methods. McGraw-Hill, 1983.

Turbines, Gas

During the twentieth century, the gas turbine was developed to fit many applications on land, sea and in the air. From early beginnings, the gas turbine came alongside, competed with, and often replaced the existing technologies of steam, water, and reciprocating internal combustion engines. Initial problems stemmed from a lack of knowledge and the techniques; the fundamentals were well enough understood, but what were lacking were the design techniques. Materials also held up developments; but after extensive experimentation, successful turbine designs were being constructed in the first ten years of the twentieth century.

The gas turbine has the advantage over traditional engines in that its combustion process is continuous and thus the equipment is less subject to cyclic heat stresses and its power is less limited—power is limited by combustion knock in spark ignition engines, but in diesel engines it is only limited by structural strength and maximum working pressures in the fuel injection systems. It also has fewer moving parts, so wear and tear is lessened. Despite these differences, the gas turbine still has the basic four functions of the four-stroke cycle but operates continuously: air is admitted, compressed, heated by burning fuel so that it expands and does work, and then the spent gases are expelled. However, unlike an ordinary engine, each of these processes takes place in a separate part of the engine and happens continuously; the oil engine has all processes within the cylinder and they follow on from each other.

Gas turbine engines can be classified according to the type of compressor used, namely radial or centrifugal flow, and axial flow. In the first of these, the air is compressed by accelerating it outward perpendicular to the longitudinal axis of the engine. Radial engines are divided into single-stage and two-stage compressors—the amount of thrust is limited because of the maximum compression ratio. Its advantages are its light weight, simplicity, and low cost.

Axial flow compressor engines may incorporate one, two or three spools (where a spool is defined as a group of compressor stages rotating at the same speed). In a two-spool engine, the two rotors operate independently of one another. The turbine assembly for the low-pressure compressor is the rear turbine unit. This set of turbines is connected to the forward low-pressure compressor by a shaft that passes through the hollow center of the high-pressure compressor and turbine drive shaft. This type of compressor can be found in the larger engines because of its ability to handle large volumes of airflow; it also has a high pressure ratio. However, it is more expensive to make, heavy in comparison with a radial engine, and

more susceptible to damage by foreign matter being drawn in—which is why a "bird strike" is so damaging for an aircraft.

Gas turbines are limited by the resistance of the turbine blades to the intense heat generated within the machine, but sufficient experimentation had been undertaken to allow work to be done commercially by turbines by the outbreak of World War II. An early commercial producer of the machines was the Swiss firm of Brown Boveri, later ABB.

Initially, the engines produced were for industrial power applications; gas turbine-powered electricity-generating sets being a main use. The transport options were also being explored for this source of power. Ships and locomotives had been built that utilized steam turbines, and gas was seen as a further possibility. Gas turbine locomotives were built using Brown Boveri units for use in Switzerland, and Britain experimented with three gas turbine locomotives, two in 1947, of which one was by Brown Boveri, and one in 1959. The concept continued in trials up until the 1970s for railways, as Britain's experimental Advanced Passenger Train was driven by gas turbines and at one stage held a speed record for a rail vehicle thus powered. The most successful use of locomotives powered by gas turbines was in the U.S., though high fuel costs and competition from diesel locomotives meant that this use was over by 1969. The problem with simple gas turbines in transport is that without heat exchangers the efficiency is poor at part load; their use in the U.S. was on a particular Southern Pacific line where there was a steady gradient on either side of a mountain pass, where the engine was run at full power to the top and then shut down for the descent.

It was not long before thoughts of gas turbines to airborne transport were conceived—as had happened and been tried in earlier years on steam and internal combustion power. The course that events then took had a direct bearing on the diminution of the world and ultimately paved the way for the globalization of markets that was the position at the close of the twentieth century. Almost all modern applications have been developed from the construction and manufacturing techniques of aeroengines. The thin sections of high-temperature metals in aeroengine construction were initially used solely for weight reduction, and people who were developing land-based engines still thought in terms of thicker sections of traditional metals, except, of course, in turbine blades and disks. Then they realized that these thin sections reduced thermal stresses and allowed quick startup and shutdown compared to "old fashioned" engines. (It is the time taken for heat to "soak" through the thick materials without excessive temperature gradients that takes so long to start up and shut down steam turbines). By the time that this was realized, jet engines and turboprops had become big enough to be useful in marine and land applications, and the number of aeroengines being manufactured made it much cheaper to adapt them for other uses than to design special engines. Among the most popular land-based turbines today are Rolls-Royce Avons and RB211s.

The development of specialized materials contributed to the progress of the thrust to weight ratio of the turbine, especially in aircraft usage. Titanium and nickel alloys have been substituted for steel, while aluminum has virtually disappeared from use in aeroengines. The future suggests that there is great potential for composites of various types, including carbon, ceramic matrix, and metal matrix composites, depending of course on achieving low cost manufacture and cost effective exploitation. The range is necessary because of the temperature variation being so great within a turbine. Materials also supply the enabling technology for improvements in performance and reliability.

Serious development of the gas turbine and its application in jet engines began around 1935 in both Britain and Germany. In Britain the name of Sir Frank Whittle will always be linked with jet engines. He wrote a paper in 1930 outlining the principles of gas turbines and jet propulsion, and took out a patent on his ideas, but due to a lack of support for practical development of his theories he let the patent lapse in 1936. Eventually with some backing, he formed Power Jets Ltd. to experiment and develop full-size engines, resulting in the first jet-powered flight in Britain, which took place on 15 May 1941.

The Germans were, however, the first to actually fly a jet-powered craft in 1939 after experimentation dating from 1935, but it is believed that their first engine to run satisfactorily did so only after Whittle's. They were therefore able to use the technology in World War II, as did the British, flying Gloster Meteor aircraft fitted with jet engines. By that time, the course of the war was well advanced and the U.S. had yet to fly a craft in combat, but after the war, gas turbines became very widespread in both civil and military aviation fields. The jet engine was developed further and it became possible to travel at previously unheard of speeds for very long distances. As time passed,

supersonic flight became possible for the first time and was offered commercially with the introduction of the Concorde aircraft in 1973.

On the maritime scene, gas turbines have worked well alongside the steam turbine. Their application on naval terms has been particularly successful, with the British Navy using only gas turbines to power its major warships since 1969. Gas turbine-powered cars have also been tried with some success, but nothing on a large commercial scale has been seen on the world's highways.

Gas turbines became an accepted form of power for transport and industry, and a journal, *The Oil Engine and Gas Turbine*, devoted its pages to them. Electric power generation has become the largest use of the turbine after air transport. The advantages of their application in this field are quick starting, the option of remote operation, low costs, and viable economics. Interest in power generation began in 1939, but the increase in oil fuel costs from 1973 alerted many to the possibilities of using the power in generating plants. This interest was also spurred by breakdowns with existing plant, often because high-level management decisions had pushed the introduction of very large sets too quickly and they were not adequately developed, resulting in delays in the commissioning of other generating stations.

The production of gas turbines continued into the twenty-first century for many of the uses outlined earlier, plus applications in powering offshore installations and compression on gas pipelines. They have not been taken as far commercially as once envisaged, due to limitations on their fuel supplies and the perceived advantages not meeting expectations. Many attempts have been made to overcome the poor efficiency at low power by using heat exchangers; these have mostly suffered from short life. However, gas turbines are an important source of power in the world, and constant development drives their improvement as a result of their use in the vast majority of large and fast aircraft in service.

See also **Turbines, Gas in Aircraft; Turbines, Gas in Land Vehicles**

ANTHONY COULLS

Further Reading

Kennedy, R. *Modern Engines and Power Generators.* Caxton Publishing, London, undated.
Kolin, I. *The Evolution of the Heat Engine*. Longman, London, 1972.
Miller, S. Advanced materials mean advanced engines. *Mater. World*, 4, 1996.
Robertson, K. *The Great Western Railway's Gas Turbines: A Myth Exposed.* Alan Sutton, Stroud, 1989.
Rolls-Royce. *Power for Air, Sea and Land.* Rolls-Royce, London, 1982 and 1989.
Whittle, F. The early history of the Whittle jet propulsion gas turbine. *Minutes Proc. Inst. Mech. Eng*., 152, 1945.

Turbines: Gas, in Aircraft

Based on the technology of the land-locked and heavy steam turbine, by the mid-1930s development began in earnest on the gas turbine for aircraft. In the end, the evolutionary development in technology had a revolutionary impact on transportation systems.

The turbojet engine for aircraft is a connected three-stage system. The compressor, in either the axial-flow or centrifugal-flow form (or a combination thereof), compresses entering air. The compressed air enters the next stage—single or multiple combustion chambers—where it is mixed with fuel and fired. The super-heated, expanded air flows over a turbine as it exits the system, producing thrust in the form of exhaust. The turbine is connected to the compressor on a rotating shaft (or shafts) that turns the entire system. The rotational forces of the turbojet system are not as strenuous as those in a conventional piston engine, and are therefore more mechanically efficient. Although the thrust in the form of exhaust is the most simplified use of the turbojet engine, the shaft horsepower can also be harnessed to drive propellers, drive gears, or applied to other uses (e.g., the 1950s American Chrysler Turbine automobile).

During the first three decades of the twentieth century, a number of developmental programs were considered around Europe and the U.S, but most were "paper projects." French, Swedish, and Swiss designers came up with ideas for turbine engines for use in aircraft, but materials, funding, and interest were practically nonexistent before the mid-1930s. By then, in the quest for speed in the air and active aerial rearmament programs, interest was rekindled in turbojet power for aircraft.

In two unrelated projects, one German and one Briton designed, built, and tested turbojet engines prior to the start of World War II. In Germany, the young physicist Dr. Hans von Ohain developed first the theory, then a working centrifugal-flow turbojet as an aircraft power plant. In Britain, a fundamentally similar, but noticeably different centrifugal-flow turbojet was designed and tested by Flight Lieutenant Frank Whittle (a Cambridge University mechanical engineering graduate) under the aegis of his new company Power Jets Inc.

Ohain's design was the first to reach fruition, with the financial backing of the German aircraft magnate Ernst Heinkel, director of the aircraft-manufacturing firm that bore his name. The first experimental turbojet aircraft, the Heinkel He-178, flew for the first time on 27 August 1939, four days before the outbreak of World War II. The British program, with Frank Whittle leading theoretical developments, was hampered by a lack of funding and materials constraints and required more time to put into operation. Whereas the Whittle WU (for Whittle Unit, his first engine) ran for the first time in April 1937, the modified W1 engine was not ready for flight-testing until 1941, when the experimental Gloster-E28/39 took off for the first time in April. Thus, the Germans had an apparent head-start in the race to develop jet engine technology.

However, the Germans during World War II lacked a coherent strategy, especially in technological development. The German emphasis on theoretical application of scientific principles encouraged them to seek out the most advanced designs possible in opposition to using practical applications. In the end, they turned to the potentially more powerful axial-flow turbojet engine to power its nascent jet force, while the British continued centrifugal-flow turbojet development. The end result was that although the Germans were able to commit jet-powered aircraft to combat operations, materials limitations and the Allied combined bomber offensive thwarted extensive production. The German equipment was revolutionary for the time, but production figures were relatively low, as was combat commitment; in addition, operations and tactics were a constant trial and error process with the revolutionary technology. The British, on the other hand, were able to commit turbojet-powered aircraft (in the form of the Gloster Meteor) to combat operations as Home Defense fighters.

During the war, two other jet programs developed directly from the German and British technology. The German's aided the Japanese, their Axis partners, in developing the Nakajima Kikka, a design inspired by the German Messerschmitt Me-262 and powered by BMW 003 axial-flow turbojets. However, the Kikka did not fly until the last days of the war and only as a prototype; it had no impact on the final outcome. The American jet program was inspired by the British Whittle developments. The Americans copied the Whittle engines directly, under license to General Electric, and used two to power the Bell XP-59a Airacomet. The Airacomet first flew on 1 October 1942. Although the Airacomet was superceded by the Lockheed P-80 Shooting Star before the end of the war, the Americans did not commit jet aircraft to combat operations.

Following the war, both military and civilian aircraft interests accepted the potential of the turbojet. Aircraft engine manufacturing firms continued development of turbojet engines for use as aircraft power plants. In addition to centrifugal-flow and axial-flow turbojets, the development of other related projects were initiated. Turboprops (turbojets that drive propellers), ramjets, and turbofans were all developed in order to explore the potential advantages of the engines with regards to speed, endurance, altitude, and efficiency.

Today, commercial aviation continues the development of high-flow bypass turbofans, which maximize efficiency in high-altitude sustained subsonic flight. Military applications include these engines, mainly for transport and bomber aircraft, while development continues on high-speed axial-flow turbojets with afterburners for fighter aircraft, to maximize speed and thrust at the expense of endurance.

Since 1937, the development of the turbojet has revolutionized air transportation. Initially conceived for military purposes with an emphasis on speed, the turbojet engine for aircraft has developed into a prime mover for various aircraft applications. Combining high efficiency and speed at high altitude, the turbojet engine has revolutionized air travel for the commercial passenger as well as the military pilot. Supersonic flight became possible for the first time and was offered commercially with the introduction of the Concorde aircraft in 1973.

See also **Civil Aircraft, Jet Driven; Turbines, Gas**

MIKE PAVELEC

Further Reading

Bathie, W. *Fundamentals of Gas Turbines*. New York, 1996.
Constant, E. *The Origins of the Turbojet Revolution*. Baltimore, 1980.
Driggs, I. *Gas Turbines for Aircraft*. New York, 1955.
General Electric Company. *Seven Decades of Progress: A Heritage of Aircraft Turbine Technology*. General Electric, Fallbrook, CA, 1979.
Gunston, B. *World Encyclopedia of Aero Engines*. Wellingborough, 1986.
Institution of Mechanical Engineers. *Development of the British Gas Turbine Jet Unit*. Institution of Mechanical Engineers, London, 1947.
Judge, A. *Gas Turbines for Aircraft*. New York, 1959.
Kerrebrock, J. *Aircraft Engines and Gas Turbines*. Cambridge, MA, 1977.
Leyes, R. *The History of North American Small Gas Turbine Aircraft Engines*. Reston, VA, 1999.

Neville, N. *Jet Propulsion Progress; the Development of Aircraft Gas Turbines*. New York, 1948.
Oates, G. *Aerothermodynamics of Gas Turbine and Rocket Propulsion*. Washington D.C., 1988.
Oates, E., Ed. *Aircraft Propulsion Systems Technology and Design*. Washington D.C., 1989.
Schlaifer R. and Heron, S. *The Development of Aircraft Engines and Fuels*. Chicago, 1950.
Treager, I. *Aircraft Gas Turbine Engine Technology*. New York, 1979.
Whittle, F. *Gas Turbine Aero-Thermodynamics: With Special Reference to Aircraft Propulsion*. Oxford, 1981.
Wilson, D. *The Design of High-Efficiency Turbomachinery and Gas Turbines*. Cambridge, MA, 1984.

Turbines: Gas, in Land Vehicles

The gas turbine has found widespread use in the aviation, marine, and stationary power areas. However the gas turbine has only seen limited use in land transportation.

As various companies began to experiment with gas turbines in the 1920s and 1930s some gave thought to using the turbine as a source of motive power for land vehicles. The turbine promised much higher power-to-weight ratios than conventional reciprocating engines and also had the capability of using cheaper fuels such as industrial heating oil, diesel fuel, and even powered coal. As with gas turbines in aviation, most development has occurred since World War II.

There have been no true production automobiles manufactured with gas turbine engines. However, there have been many experimental vehicles. Some of these vehicles utilized aviation engines and constructed vehicles around them. The exhaust blast was utilized for propulsion as in jet aircraft. These vehicles were not designed for general highway travel and were exclusively confined to high-speed, often record breaking, runs. Many individuals and groups have built exhaust-powered gas turbine vehicles for drag racing, automobile shows, and land speed record attempts. A more practical way of utilizing gas turbines to power a vehicle, tried by several manufacturers, was to connect the turbine to shafts for wheel-driven automobiles. Generally hydrostatic automatic transmissions were used to control the speed of the vehicle. Among the most successful of the gas turbine automobiles were Chrysler's several efforts but many other manufacturers, including Ford, General Motors, Rover, Volvo, Nissan, Toyota, Volkswagen, and Daimler Benz experimented with wheel-driven gas turbine automobiles. Rover produced the first true gas turbine-powered, wheel-driven automobile in 1950. Chrysler had been experimenting with turbines since the early 1950s and achieved its largest success in 1963–1964, when it produced 50 vehicles and distributed them on a test basis to typical consumers all over the U.S. General Motors experimented with geared gas turbines in heavy trucks and buses in addition to automobiles. Most of the turbines used in these vehicles were single-stage, centrifugal compressor designs. Some incorporated heat regenerators. While the vehicles worked fairly well, there were many issues to address. Fuel consumption was generally worse than conventional gasoline or diesel piston engines. There was also a lag in acceleration experienced in many of the test automobiles due to the spool-up time, or the time to reach operating speed, of the turbine. The consequences of catastrophic failure worried designers, and emissions were also a concern after the clean air act and the creation of the EPA in the United States. Despite these concerns, experimentation continued throughout the 1970s and 1980s.

While unsuccessful in civilian automotive use, the gas turbine was put to good use in the U.S. Army's M-1 Abrams main battle tank, built by General Dynamics and first delivered in 1980. The high power-to-weight ratio of the turbine as well as its smaller size and the capability for multiple fuels were the primary reasons that turbine power was chosen for the M-1. Many of the concerns raised in automobile testing, such as fuel consumption and emissions, were deemed to be of secondary importance in a military application. The M-1 has a 1120-kilowatt gas turbine coupled to a six-speed automatic hydrokinetic transmission.

The most successful gas turbine use in ground transportation has been in the railway arena. Early experiments began prior to World War II with full-scale locomotives entering production during and shortly after the war. The Swiss firm Brown-Boveri built a 1640 kilowatt locomotive for the Swiss federal railway in 1941. British Rail rostered three experimental locomotives during the 1950s and early 1960s. The first two, delivered in 1950 and 1952 respectively, were gas turbine electrics while the third, built in 1960, was a direct-drive geared locomotive. They were used mostly in the western region and were eventually withdrawn from service partly due to high fuel costs. In the U.S., both Westinghouse and General Electric (GE) adapted aviation-type gas turbines to railroad use. Westinghouse built one test locomotive in 1950 but did not follow with a full line of locomotives. General Electric was more successful. It built one test locomotive in 1948 for the Union Pacific railroad and followed with several different production models manufactured until 1961. There

were three distinct types of gas turbine locomotive in service on the Union Pacific. The first was the single GE product of 1948. The second were a group of 24 3360-kilowatt GE turbines produced from 1952 to 1954. These were the first truly operational gas turbine locomotives in the U.S. The final group of gas turbines on the Union Pacific were the monstrous 6340-kilowatt "Big Blow" turbines built from 1958–1961. These were the largest and most powerful single locomotives in the world when produced. The Union Pacific was the only North American railroad to use gas turbine locomotives in revenue service. Both the Westinghouse and GE locomotives were gas turbine electrics. The multistage, axial flow gas turbine drove an electric generator, which powered traction motors located on the axles. Both were designed to burn bunker C oil, a heavy industrial fuel. The Union Pacific also experimented with a coal-burning gas turbine locomotive. This turbine had many more difficulties than the earlier turbines and was purely experimental. The Union Pacific gas turbines were very loud, and could only be operated in rural and isolated areas. Because of these concerns and the rising cost of fuel, all were withdrawn from service by the end of 1971.

Apart from locomotives, railways have also made use of gas turbines in self-propelled trains and railcars. SNCF, the French national railway, achieved success in the late 1960s and 1970s with the RTG gas turbine-powered high-speed train. Amtrak imported six in 1973 for use in the U.S. and several more were manufactured under license by the Rohr Corporation. The United Aircraft Corporation turbine train was used in both the U.S. and Canada in the 1970s and early 1980s. The British effort at a high-speed gas turbine train, the APT or Advanced Passenger Train, was not successful.

See also **Rail, Electric; Rail, High Speed; Tanks; Turbines, Gas**

JEFF SCHRAMM

Further Reading

Advanced Gas Turbine Systems for Automobiles. Society of Automotive Engineers, Warrendale, PA, 1980.

Alternative Sources of Railroad Motive Power. Association of American Railroads, Washington, D.C., 1985.

Carrol, W. *Automotive Gas Turbines*. Coda Publications, Studio City, CA, 1963.

History of Chrysler Corporation Gas Turbine Vehicles. Chrysler Corp., Detroit, MI, 1964.

Lee, T.R. *Turbines Westward*. T. Lee Publications, Manhattan, KS, 1975.

Simmons, J. and Biddle, G., Eds. *The Oxford Companion to British Railway History*. Oxford University Press, Oxford, 1997.

U.S. Passenger Rail Technologies. U.S. Congress Office of Technology Assessment, Washington D.C., 1983.

Turbines, Steam

The first steam turbine, of which there is any record, was made by Hero of Alexandria more than 2000 years ago. This simply demonstrated that a jet of steam, impinging on a paddle wheel, could convert heat energy into mechanical energy. In the late nineteenth century significant improvements in the efficiency of conversion were made by, among others, Sir Charles Parsons on Tyneside, U.K. and Charles G. Curtis in the U.S.

Early steam engines up to that time had involved very high rotational speed, which was difficult to utilize for many purposes unless speed-reducing gearboxes were employed. Parsons had deduced that moderate surface velocities and speeds of rotation were essential if the "turbine motor" was to receive general acceptance as a prime mover. His early designs arranged to divide the fall in pressure of the steam into small fractional expansions over a large number of turbine wheels in series so that the velocity of the steam over each wheel was not excessive.

At the close of the nineteenth century, many local power stations employed reciprocating steam engines to drive electric generators. Steam turbines had the advantage over reciprocating steam engines, which were based on the movement of a piston in a cylinder, of being lighter and more efficient. The Curtis multiple-stage steam turbine (patented in 1896, sold rights to General Electric in 1901) occupied a smaller space and cost much less than contemporary reciprocating steam engine-driven generators of the same output. The Curtis turbine was also shorter than the Parsons turbine, and was thus less susceptible to distortion of the central shaft.

The work that Curtis, Parsons, and others carried out in the development of steam turbines allowed large central power stations to be developed, providing electricity for the growing demand during the early 1900s. Early machines at the beginning of the twentieth century had ratings typically of less than 1 megawatt.

Operating Principle

Steam is supplied to a turbine at high pressure and temperature from a boiler and the energy in the

steam is converted into mechanical work by expansion through the turbine. This expansion takes place through a series of fixed blades or "nozzles" that direct the steam flow onto moving blades on a turbine disk, which rotates under the action of the steam to produce mechanical power in the rotor shaft. A row of fixed blades and its associated moving blades are known as a "turbine stage." Many such stages in series form a turbine cylinder. The fixed blades are attached to the turbine casing, which contains the steam pressure. In large output turbines, the steam flow and expansion ratio are too great to be contained within a single turbine rotor and casing and several such turbine cylinders are combined to achieve the necessary output.

Steam turbine development in the U.K. progressed from 30-megawatt sets at the end of the 1930s, operating at 450°C and 41 bar, through to 100-megawatt sets in the mid-1950s, to 660-megawatt sets operating at 158 bar and 565°C in the late 1970s. Designs for 1300-megawatt electric sets existed at the end of the twentieth century.

Applications

The chief applications for steam turbines are:

- Electricity generation with machines ranging in size from comparatively low powers of the order of hundreds of kilowatts up to about 1300 megawatts in central power stations. Electric power generation on land is now produced almost exclusively by steam-driven turbine generators.
- As part of an industrial process to produce electricity but where the steam is used for other purposes after its exhaust from the turbine.
- Variable speed drives, such as boiler feed pumps and blast-furnace blower drives.
- Marine propulsion, such as large ocean liners of the early twentieth century. Parson's demonstration of *Turbinia* at the 1897 Spithead Naval Review was the world's first steam turbine-driven ship. Now largely superseded in conventional shipping, steam turbines are still used to drive nuclear-powered vessels.

Characteristics and Efficiency

The output characteristics of a steam turbine are based on the so-called "Rankine" cycle. Water is pumped into a boiler in which heat is supplied to convert the water into steam. This steam is then utilized in a steam turbine to do work (e.g., to drive an electric generator). The steam exhausting from the turbine is condensed and then pumped back into the boiler to complete the cycle.

The output of any heat engine such as a steam turbine is governed by the simple relationship that defines its efficiency:

$$\text{Efficiency}_{\text{max}} = 1 - (T_{\text{out}}/T_{\text{in}})$$

where T is the temperature of the steam (expressed in degrees Kelvin). This gives the so-called "Carnot efficiency," which is an ideal that cannot be achieved in practice, only approached. It will be seen that raising the input temperature of the steam as high as possible and reducing the temperature at which the steam is condensed back to water as low as possible are essential to achieve high heat-to-power efficiencies.

Efficiency Improvements

Improvements in the overall cycle efficiency are achieved by design refinements such as bled-steam feed heating, where some of the steam passing through the turbine is extracted to preheat the water entering the boiler. The improvement in efficiency obtained with feed-water heating increases with turbine inlet pressure.

Another refinement is "reheat" where the steam passing between the different stages of the turbine is raised in temperature by circulating it back through a section of the boiler. Increases in efficiency of the order of 3 to 5 percent can be gained in this way although, as operating steam temperatures are raised, there is a progressive reduction in the improvements that can be gained in this way.

Economies of Scale

In postwar years the trend toward larger and larger units in power stations enabled marked reductions in capital cost per kilowatt of capacity. It also enabled improvements in thermal efficiency to be made, partly as a direct result of the increases in size, and partly because the economic steam pressure increases with size. Steam temperatures have increased also over the period up to 565 to 593°C which is near the limit for ferritic steels in the turbines.

The highest efficiency steam turbines at the end of the twentieth century were to be found in Japan

where improvements in manufacturing technology and materials, together with blades designed utilizing computer-based three-dimensional flow analysis and "super-critical" steam conditions of 250 bar, 600 to 610°C, have resulted in 1050-megawatt machines giving overall power station efficiencies of 49 percent.

See also **Electrical Energy Generation and Supply, Large Scale**

IAN BURDON

Further Reading

Hills, R.L. *Power from Steam: A History of the Stationary Steam Engine*. Cambridge University Press, Cambridge, 1993.

Ramage, J. *Energy: A Guidebook*, 2nd edn. Oxford University Press, Oxford, 1997, chapters 5–7

Useful Websites

Parsons, C.A. *The Steam Turbine*. Rede Lecture. Cambridge University Press, 1911: http://www.history.rochester.edu/steam/parsons/

U

Ultracentrifuge

As a scientific instrument, the ultracentrifuge is used primarily in biochemical research to sediment substances from a solution. A rotor with containers holding the solution is spun at high speeds, causing the solute to move towards the periphery of the container under influence of the generated gravitational fields. Historically, however, ultracentrifuges were first developed for analytical purposes to determine the size of colloidal particles.

Early High-Tech Machine

Beginning about 1910 and through the 1920s, Swedish chemist Theodor Svedberg analyzed the particle sizes of colloids. Under the influence of gravity, these particles gradually separated from a solvent at a speed depending upon their size. To improve the accuracy of his method, he began to use centrifuges to generate forces many times that of the earth's gravitational field. He developed this idea into a practical method and could subject the colloid particles to forces up to 5000 g (5000 times the force of gravity). In analogy with the common methods of ultrafiltration and the ultramicroscope, he called the apparatus an ultracentrifuge.

Svedberg used the new apparatus to determine particle sizes of both inorganic and organic colloids including proteins, expecting that these, like inorganic colloids, would show a wide distribution of particle sizes. However, when he analyzed hemoglobin, all particles seemed to have the same size. This suggested that this protein could be a well-defined molecule, which was a revolutionary thought at the time. To be sure, however, further analyses would be needed using fields of the order of 100,000 g.

This 20-fold improvement was difficult to achieve, but he succeeded after investing huge sums of money and conquering a variety of problems. The rotor, driven by oil turbines to facilitate lubrication of the bearings, spun in a hydrogen atmosphere to reduce heat production. The substance to be analyzed was placed in a special container and the sedimentation of the particles caused by the gravitational forces was recorded by illuminating the process with a special light source and taking photos at regular intervals. Some hours of calculation were needed to deduce the particle sizes from these photographs.

After experiments with this new ultracentrifuge Svedberg concluded that hemoglobin was indeed a monodisperse protein. After this surprising result, research at his laboratory became almost exclusively focused on proteins. In the following decade he tried to improve his ultracentrifuges, especially to increase the forces that could be generated. Almost all parts of the apparatus were optimized, including the shape of the turbines and turbine chambers, oil inlets, bearings, type of oil used, rotor balancing method, rotor size, and so on.

By 1937 Svedberg considered that he had reached the limit of modifications. He used his latest ultracentrifuge, which generated 400,000 g, until his retirement in 1949, after which the protein research was continued by his former colleagues. It was not until the mid-1970s, half a century after the first apparatus was developed, that the oil-turbine ultracentrifuge was taken out of use.

Vacuum Ultracentrifuge

American physicist Jesse Wakefield Beams studied optical phenomena in the 1920s. In his research he used rapidly rotating mirrors mounted on small conically shaped spinning tops (of the order of 1–2 cm in diameter) that were driven by compressed air. After 1930 he started to make these tops hollow, which allowed him to use them as small centrifuges. Because of the high speeds that could be achieved, Beams called them ultracentrifuges. He identified a wide variety of applications for the apparatus, including Svedberg's method of determining molecular weights.

Edward Greydon Pickels, one of Beams' students, developed the apparatus further for this application. This proved difficult because the high speeds heated the rotor and caused convection currents in the sedimentary solution. Pickels tried various solutions until in 1935 he produced a design in which the rotor spun in vacuum. A small wire, passing through a vacuum-tight gland, connected the rotor to the driving air turbine. This design solved all convection problems, allowing forces up to 1,000,000 g.

This design attracted the attention of scientists at the Rockefeller Institute for Medical Research in New York, and Pickels developed two types of ultracentrifuge further for them. The first type, the analytical ultracentrifuge, was used to determine particle sizes and was analogous to Svedberg's method. The second type, the preparative ultracentrifuge, could separate a substance from a solvent and was primarily used for concentrating viruses. Beams attempted to make his apparatuses as simple as possible to allow many scientists to use them. His instrument makers made some ultracentrifuges for others, which led to a limited distribution in the scientific world. Around 1937 his vacuum ultracentrifuge was marketed by an American company, but this resulted in commercial failure.

Svedberg was also prepared to sell his ultracentrifuges to others, but the apparatus was extremely expensive—in the order of \$20 000—which was an enormous amount of money for scientists in those days. In the early 1940s the apparatus was marketed by a Stockholm company, but once again it was a commercial failure.

In 1946 Pickels was approached by a salesman who wanted to market an analytical ultracentrifuge based on Pickels' design. Together they formed Spinco, or Specialized Instruments Corporation. Pickels considered his design at the time too complicated and developed a more easily operated, foolproof version. Sales, however, remained low, and Spinco nearly went bankrupt. Subsequently, Pickels concentrated on developing the preparative ultracentrifuge, which seemed to sell reasonably well. This gave Spinco sufficient financial power to continue the production of the analytical ultracentrifuge as well, although in small numbers. Over the years, that number gradually rose.

Increasing numbers of scientists used ultracentrifuges for biochemical research, forming a research community that met at symposia and conferences at regular intervals. The ultracentrifuge was still not fully developed. As more scientists used them, each with their own approach and interest, new desires and problems in connection with the apparatus were articulated. Over the years a variety of ultracentrifuges have been developed to the point that it became a very common instrument in laboratories for many types of biochemical research.

BOELIE ELZEN

Further Reading

Beams, J.W. High speed centrifuging. *Rev. Mod. Phys.*, 10, 245–263, 1938.

Elzen, B. *Scientists and Rotors: The Development of Biochemical Ultracentrifuges*. Dissertation, University of Twente, Enschede, 1988.

Elzen, B. The failure of a successful artifact—the Svedberg ultracentrifuge, in *Center on the Periphery: Historical Aspects of 20th-Century Swedish Physics*, Lindqvist, S., Ed. Science History Publications, Canton, MA, 1993, pp. 347–377.

Pickels, E.G. High-speed centrifugation, in *Colloid Chemistry*, vol. 5, Alexander, J., Ed. Chemical Catalog Company, New York, 1944, pp. 411–434.

Rånby, B. Physical chemistry of colloids and macromolecules. *Proceedings of the Svedberg Symp*osium, Blackwell, London, 1987.

Schachman, H.K. *Ultracentrifugation in Biochemistry*. Academic Press, New York, 1959.

Svedberg, T. and Pedersen, K.O. *The Ultracentrifuge*. Clarendon Press, Oxford, 1940. Reprinted by The Johnson Reprint Corporation, New York, 1959. German edition: *Die Ultrazentrifuge—Theorie, Konstruktion und Ergebnisse*. Verlag von Theodor Steinkopff, Dresden, 1940.

Ultrasonography in Medicine

The word *ultrasound* refers to sound waves beyond the range of human hearing (oscillations above 20,000 cycles per second or 20 kHz). The dog whistle is an often-used example. Both sound and ultrasound waves are mechanical vibrations that can only propagate through matter (liquids, solids, and to some extent, gases). The composition and temperature of that matter determines the velocity

of these waves. Whenever these waves encounter a boundary (such as an organ wall) or interface between two substances, there is a decrease in velocity, and some of the energy is reflected back as an echo, while the rest of the energy passes through to the next interface.

In clinical ultrasonography, a transmitter produces a short pulse (typically a few millionths of a second) of high-frequency electrical oscillations (1 to 10 MHz, or million cycles per second). The transducer, acting like a loudspeaker, converts this electric signal into a pulse of mechanical vibrations of about the same frequency and duration. With the transducer pressed firmly against the patient's body, and acoustically coupled to it with gel or oil, the pulse of ultrasound energy enters with little reflection at the skin and propagates inward through soft tissues and fluids. Then, quiescent and acting as a microphone, the same (or another) transducer senses any reflected, much weaker pulses of ultrasound, and transforms them back into electrical signals. Echoes from distinct organs, blood vessels, and other structures are amplified and processed by the receiver, and are sent to a computer, which keeps track of their return times and amplitudes. About a thousandth of a second later, another pulse is produced and sent off in a slightly different direction through the body, and the whole process begins anew. From echo data generated in this fashion, the computer can create a real-time image in which one sees the arms and legs of a fetus move around or a heart valve open and close.

Ultrasound is particularly useful in the study of soft tissues and organs that are too similar to provide adequate x-ray image contrast. Doppler ultrasound can detect and monitor the flow (or lack thereof) of blood in the arteries and veins. Ultrasound does not use ionizing radiation and usually costs less than other imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI).

Ultrasonics originated in the domain of physics. The basic properties of sound were described in the classic work of Lord Rayleigh on *The Theory of Sound* (1877). The discovery of the piezoelectric effect by the French physicists Pierre and Jacques Curie in 1880 provided a way for generating ultrasound waves and became the principle of ultrasonic transducers. They found that certain ceramic materials, such as quartz, could convert electric signals to acoustic signals, such as ultrasound waves, and vice versa. The most common transducers are made of ceramic materials that have been processed to have these piezoelectric characteristics. It was neither possible to use this discovery for generating high-frequency sound for many years in the absence of suitable oscillators, nor to detect the low-amplitude electrical signals generated without suitable amplifiers.

While the early medical applications of ultrasound were diagnostic, those during the interwar years were therapeutic and very conventional. During the 1930s ultrasound was used for therapeutic heating, particularly in physical therapy, for the sterilization of biological preparations such as vaccines, and for cancer therapy, often in combination with x-rays. Some researchers during the late 1930s and early 1940s studied the biologic effects of ultrasound, and a few, such as the French physiotherapist Andre Dernier and the brothers Dussik (neurologist Karl and physicist Friedrich), suggested that ultrasound could be used to produce images of interior body structures such as the brain. However, the development of diagnostic medical imaging using ultrasound had to await the end of World War II and the postwar demobilization of expertise before the military sonar and radar techniques based on echo principles could be more fully applied to medical uses.

The emergence and development of diagnostic medical ultrasound was a worldwide phenomenon. During the decade 1948–1958, researchers in Japan, Europe, and North and South America worked mostly independently, with very little collaboration or exchange of information, to adapt military and industrial ultrasonic equipment to medical uses. In Japan, pioneering work was done by the physicist Rokuro Uchida at the research laboratories of Nihon Musen Company (now Aloka) and by physicians Kenji Tanaka and Toshio Wagai at Juntendo University School of Medicine in Tokyo. Using flaw detectors adapted for medical ultrasonic use they scanned the brain (to detect intracerebral hematomas and tumors), the gall bladder (to detect stones), and the breast (to detect tumors). These early machines displayed information (returning echoes of ultrasonic pulses) in A-mode (one-dimensional) as bright dots on an oscilloscope screen corresponding to points within the body. Another important group in Japan formed around Shigeo Satomura and Yasuharu Nimura at Osaka University. They discovered that the Doppler effect could be applied to ultrasonic energy and used it in their cardiovascular investigations, particularly of heart valves and blood flow, publishing their first results in 1956.

In the U.S., important early work in medical ultrasound was done by the internist George Ludwig, first at the Naval Medical Research

Institute in Bethesda, Maryland and then with the neurosurgeon Thomas Ballantine Jr. at the Massachusetts General Hospital and a group of physicists at the Massachusetts Institute of Technology's Acoustics Research Laboratory; by the surgeon John Wild and engineer John Reid in Minneapolis, Minnesota; by the radiologist Douglass Howry and nephrologist Joseph Holmes in Denver, Colorado; and, by the physicist William Fry and his group at the University of Illinois at Champaign.

Ludwig, using exclusively A-mode presentation, studied the velocity of ultrasound in various animal tissues, which became standards for later researchers, and worked on the detection of gallstones and foreign bodies embedded in tissues—in principle a flaw detection approach.

Wild began his work in 1949 using a Navy radar trainer operating at 15 MHz to measure the thickness of excised bowel tissue. The discovery that echoes from tumor-invaded tissue could be distinguished from those produced by normal tissue in the same sample led him to apply ultrasound to cancer detection, particularly of the breast. Together with Reid he built a B-mode contact scanner, which provided a cross-sectional, two-dimensional picture of the plane of the body scanned (and thus more accurate position information than A-mode). This permitted real-time scanning so that images appeared directly on the screen during the examination with no need for intervening film development.

Unlike Wild whose focus was "tissue characterization," Howry was primarily interested in producing accurate cross-sectional anatomical images as the basis of medical diagnosis. Using surplus U.S. Air Force radar equipment, he built a pulse-echo electronic scanner in his basement in 1949. In 1950 he recorded his first cross-sectional pictures obtained with ultrasound using a 35 mm camera. However, with only a horizontal scanning motion, it was not possible to make the kind of accurate anatomical pictures of living tissue he wanted. However, by 1951, working with engineers Roderick Bliss and Gerald Posakony, Howry built scanners that utilized a cattle watering tank and later the rotating ring gear from a B-29 gun turret as a water immersion tank system to introduce multiposition, or compound scanning, which resulted in the removal of "false" echoes and therefore better images. While one motor moved the transducer around the patient, another provided a second back-and-forth motion, resulting in compound scanning of the immersed subject. The problems inherent in this kind of water-bath coupling system for ill patients were obvious. Other engineers worked with Howry to build a portable scanner that utilized the principle of compound scanning, came into direct contact with the skin, and had an articulated arm to which the transducer was attached and could be easily manipulated and moved. This "Porta-arm" scanner, marketed by Physionics, enjoyed widespread clinical use beginning in 1964.

The work of Fry and his group was in the older tradition of ultrasound use in biophysical investigations and therapeutics, as opposed to diagnostic applications and soft tissue visualization. During World War II and until 1946, Fry worked on the design of piezoelectric transducers at the Naval Research Laboratory in Washington D.C. After establishing the Bioacoustics Research Laboratory at the University of Illinois in Champaign, he concentrated on the use of high-intensity ultrasound as a noninvasive surgical technique to treat neurological or other brain-related disorders such as Parkinson's disease and brain tumors.

The earliest diagnostic applications of ultrasound were made in the specialties of neurology, cardiology, gynecology, obstetrics, ophthalmology, and internal medicine. They were made in Europe, the U.S., Canada, Australia, and Japan. They were made by physicians, who could identify clinical needs, working closely with engineers and physicists, who could provide technical design and engineering skills. For example, in 1953, Inge Edler, a physician at the University of Lund in Sweden, began a collaboration with physicist C. Hellmuth Hertz that eventually launched clinical echocardiography using the pulse-echo technique. At the University of Glasgow's Department of Midwifery in the mid-1950s, physicians Ian Donald and John MacVicar worked with engineer Tom G. Brown and physicist Tom Duggan to introduce ultrasound as a diagnostic modality in obstetrics and gynecology. In the U.S., the physician Gilbert Baum worked with engineer Ivan Greenwood to pioneer the use of ultrasound in ophthalmology.

Steady advances in ultrasound technology occurred in the last half of the twentieth century. The A-mode and B-mode scanners of the 1950s and 1960s gave way in the 1970s to mechanical sector, linear, and phased array scanners, and to grayscale static imaging and real-time. In the 1980s real-time, two-dimensional ultrasound was standard, and computers were harnessed to produce even higher quality images. By the late 1980s color Doppler had been introduced. In the 1990s, the greatest advances were the development of three-dimensional grayscale and color Doppler and the

introduction of ultrasound contrast agents to significantly improve diagnostic capabilities. Some of the major manufacturers of ultrasound machines have been: Acuson, Aloka, Diasonics, General Electric, Hewlett-Packard, Kretz, Medison, Philips, Picker, Siemens, and Toshiba.

Almost 25 percent of all imaging studies worldwide are ultrasound. Because of the wide accessibility and utility of ultrasonography, the World Health Organization now recommends its use after basic x-ray rather than more advanced imaging procedures such as CT and MRI.

See also **Nuclear Magnetic Resonance (NMR) and Magnetic Resonance Imaging (MRI); Tomography in Medicine**

RAMUNAS KONDRATAS

Further Reading

Blume, S.S. *Insight and Industry: On the Dynamics of Technological Change in Medicine*. MIT Press, Cambridge, MA, 1992. Chapter 3 deals with diagnostic ultrasound.

Hackmann, W. *Seek and Strike: Sonar, Anti-Submarine Warfare, and the Royal Navy*. Her Majesty's Stationery Office, London, 1984.

Koch, E.B. In the image of science? negotiating the development of diagnostic ultrasound in the cultures of surgery and radiology. *Technol. Cult.*, 34 858–893, 1993.

Tansey, E.M. and Christie, D.A., Eds. *Looking at the Unborn: Historical Aspects of Obstetric Ultrasound,* vol. 5 of the Wellcome Witnesses to Twentieth Century Medicine series. Wellcome Trust, London, 2000.

Wolbarst, A.B. *Looking Within: How X-Ray, CT, MRI, Ultrasound, and Other Medical Images Are Created and How They Help Physicians Save Lives*. University of California Press, Berkeley, 1999. Serious but non-specialized descriptions of how medical images are created. Chapter 7 deals with ultrasound.

White, D.N. Neurosonology pioneers. *Ultrasound Med. Biol.*, 14, 541–561, 1988.

Yoxen, E. Seeing with Sound: A Study of the Development of Medical Images, in *New Directions in the Sociology and History of Technology*, Bijker, W.E., Hughes, T.P. and Pinch, T., Eds. MIT Press, Cambridge, MA, 1987.

Urban Transportation

All the forms of urban transit popular in the twentieth century—buses, streetcars or trams, trolleybuses, cable cars, railroads, and subways—had their origins in the nineteenth century or before. In that century they merely served towns and cities, but in the twentieth century, enhanced mainly by electrical technology, they helped to create new and enlarged communities.

Buses and streetcars were a familiar sight in the nineteenth century. Buses developed from coaches and had their origin in Paris. There, in 1823, Stanislas Baudry, the owner of hot baths in a Nantes suburb, began running his horse-drawn vehicles to them from a stand outside a shop belonging to an M. Omnes, whose business had the slogan "Omnes omnibus"—Omnes for all. Baudry soon found people using his carriages to go to other places, so he renamed them "Omnibuses" and introduced other routes. George Shillibeer spread their use to England. He began a regular timetabled omnibus service in London on 4 July 1829. The fare was half that of a stagecoach, advanced ticket booking was unnecessary, and the omnibuses stopped wherever and whenever requested. Omnibus services were introduced across Europe and in the U.S., especially after the coming of railways.

Streetcars were pioneered in South Wales but developed in the U.S. in the late 1820s. Their use spread to Europe by way of England in the early 1860s. The lower friction between a metal wheel and track meant that a horse could pull over seven times the load of a horse pulling a bus, and new rails offered a much smoother ride than the poor roads. Thus it was to the streetcar that technology was applied, first to assist, then to supplant, the horse. Many forms of mechanical traction were tried, but only two succeeded—steam and cable—with only the latter surviving significantly into the twentieth century.

Cable Cars

Andrew Hallidie, an English-born wire rope manufacturer living in San Francisco in the late 1860s, largely devised cable car technology. After witnessing the cruel treatment of streetcar horses struggling up one of that city's famous hills, he devised a horse-free system using cable haulage. His first line ran up Clay Street and opened on 1 September 1873. It was 1584 meters long and rose 94 meters in that distance. A double track was laid. A steel wire rope of 3,353 meters was carried in an iron tube and supported on pulley wheels placed at 12-meter intervals. The cable was wound on a drum turned by a steam engine. The cars were powered by a constantly moving, endless loop of wire rope that traveled at a constant speed of 10 kilometers per hour. A grip mechanism protruded down from the car through a slot in between the running rails and into a second slot in the top of the tube carrying the cable. When the driver turned a hand wheel, the jaws of the grip opened or closed. In the closed

position its jaws gripped onto the cable, and the car moved forward. Armed with only this gripper mechanism and a handbrake, the driver had to control his cable streetcar.

Worldwide, 81 towns and cities had cable streetcars. Their use was suited to places with hills, and they found most favor in the U.S. where 66 locations in 18 states built at least one line. There were also nine lines built in the U.K. (Birmingham, Douglas, Edinburgh, Glasgow, Liverpool, Llandudno, two in London, and Matlock), one in France (Paris), two in Portugal (Lisbon), two in Australia (Melbourne and Sydney), and four in New Zealand (Dunedin). However, the expense of extending existing lines or laying down new ones meant that cable streetcars were rarely modernized, as demonstrated by the fact that in those few places where they do survive, they are period pieces or curiosities of urban transit history, maintained for tourist purposes. Also, the technology did not allow cars from one route to pass over another one, each being a separate line. Thus the routes were inflexible, and it was their passengers who had to change, from line to line and car to car. As towns and cities developed rapidly in the early part of the twentieth century, they quickly outgrew their cable cars. As technology improved the electric streetcar, cable cars even lost their advantage for climbing hills.

Almost two thirds (41) of the U.S. cable cars had closed by the end of the nineteenth century, and there was only one new line opened anywhere in the world during the twentieth century—a route extension in Dunedin, New Zealand, which only lasted three years. Most of the remainder had closed by World War II, including the 41 kilometers of lines in Edinburgh, which ended on 23 June 1923, and the massive 75 kilometers of lines in Melbourne, which used 1200 cars and last ran on 26 October 1940. Both of these systems were replaced by electric traction.

Electric Streetcars

Experiments with electric traction on railways began in the 1830s, but it was not applied to urban transit until the late 1870s. On 31 May 1879 Dr Werner von Siemens demonstrated the first practicable application of electricity to the mass movement of people with a 320-meter line at the Berlin Industrial Exhibition. A small four-wheeled locomotive drew 150 volts from a third rail to power an electric motor, the current returning via the running rails. It hauled three passenger cars, each seating six people. Over the course of the exhibition, this carried over 80,000 passengers. On 12 May 1881, von Siemens opened a more permanent electric streetcar, linking the railway station at Lichterfelde outside Berlin with its Cadet School, some 2.5 kilometers, this time using only the running rails to conduct the 100 volt power. Other streetcars employing electrical conduct through their running or auxiliary rails opened in Ireland and the U.K. in the early 1880s, while in the U.S. there were experiments being made to conduct electricity thorough overhead wires. Leo Daft, John C Henry, Charles van Depoele, and others all used twin wire systems: one positive, one negative, along which ran a small four-wheeled ‘‘troller’’—resembling a basic roller skate—attached to the streetcar by a cable. This system worked, but was not without its problems. The trollers were prone to come off the wires, and did not stop when the streetcar they were attached to did!

Frank Sprague rescued electric streetcars from becoming technological curiosities. Bound by a ridiculously tight contract to build an electric streetcar system in Richmond, Virginia, across steeply graded land over which trollers were useless, Sprague devised and perfected a single wire system, with a counter-sprung trolley pole and wheel, which conducted electricity through the streetcar, returning it via the track. He also refined the design of the electric motors and control equipment required and devised a means of mounting the motors in the streetcar's truck so that their performance was unaffected by rail joints and track undulations. When his Richmond system opened on 2 February 1888, Sprague had built the prototype for the majority of streetcar systems that would follow anywhere in the world. In essence it worked like this: the streetcar completed an electric circuit between the positive overhead wire and the negative track. Current from the wire passed down a cable in the trolley pole through a circuit breaker to a controller. This housed a series of electrical resistances, switched in and out of circuit by the action of the streetcar driver rotating a handle to preset, notched positions, which progressively allowed more and more current through. From the first few notches the current passed to the motors in series—one after the other—which produced the maximum torque to move the streetcar from stationary. Once moving, higher notches passed the current to the motors in parallel—at the same time—which kept them moving at the same speed. Advances in technology improved the performance of electric streetcars but this basic principle of their operation was unchanged.

Few technologies spread faster than the electric streetcar in the late nineteenth and early twentieth centuries. Their speed and power made them ideal for moving large numbers of people quickly. Streetcars could be loaded well beyond their stated capacity without any loss of performance, and they could be run in safety more closely together than trains. Few major towns and cities failed to have them, and some, such as Liverpool, planned new routes along with the building of new housing estates. In the U.S., electric streetcars were also built to link towns and cities. Called "interurbans," these were built in two main waves from 1901 to 1904 and from 1905 to 1908. By 1916 a maximum of 25,074 kilometers of lines had been built. Ohio had the most with 4503 kilometers; in the 1910s a journey of 1749 kilometers was possible using connected interurban systems.

Despite their obvious advantages, electric streetcars also had their drawbacks. The urban growth they fostered sometimes outpaced them, either requiring costly new routes and extensions, or providing a toehold for motorbuses. Streetcar rails also wore out, requiring constant repair and needing replacement every 15 to 20 years. In the U.K., crippling legislation placed huge financial burdens on operators for the upkeep of the roads, while in the U.S. a company called National City Lines was formed in 1936 to acquire local systems throughout the country and convert them to motorized bus operations. In 1949 it was revealed that National City Lines had ties with General Motors, Firestone, Standard Oil, and Phillips Petroleum, companies that had an interest in supplying buses, tires, and fuel. Elsewhere streetcar systems have proven remarkably resilient, and their value has been underscored by the fact that many European systems devastated during World War II were effectively rebuilt afterward. Others brought to the brink of closure, such as that in Melbourne, have been reequipped and given a secure future.

Motorbuses

Many applications were tried for the internal combustion engine in the late 1890s, including powering buses. The first regular service of gasoline-engine buses was introduced in London on 9 October 1899, using German Daimler engines. Seating capacities were low on these early motorbuses, typically 14 to 16 people, but the Peckham–Oxford Circus service, introduced on 30 September 1904 by Thomas Tilling, used 34-seater Milnes–Daimler buses. In 1905 the first motorbus services were introduced in Paris and on Fifth Avenue in New York, and they were introduced on lightly trafficked routes in Vienna in 1907. A measure of the success of these early motorbuses is given by figures for those in use in London: in 1905, 20 were in use; by 1908, 1066 were in use. The development of larger and more powerful engines allowed the seating capacity of motorbuses to rise, such that by 1919 the London General "K" type seated 46, and by 1926 Daimler–Benz of Germany produced a 60-seater double-decker, one of the first buses capable of taking a streetcar's load of passengers.

At first motorbuses were used on lightly trafficked routes, where there were insufficient passengers to justify the building of a streetcar line. They were also used to extend streetcar routes beyond their termini, to serve ever-growing suburbs. Once the seating capacity approached that of streetcars, buses posed a direct threat to the latter; and the introduction of pneumatic tires, improved suspension, and padded seats gave superior passenger comfort with which increasingly aged streetcar fleets could not compete. Finally, the wider availability of diesel engines from the late 1920s offered bus manufacturers and operators greater power at reduced cost, as they ran on a heavy oil produced at an early stage in the refining process, which was cheaply produced. Diesel engines were first used in buses in the U.K. in Sheffield in 1930 and very widely in the U.K. and elsewhere after 1933.

Trolleybuses

From 29 April 1882, streetcar pioneer Dr. Werner von Siemens began experimental operation of his 'Electromote', a four-seat, four-wheel dogcart electrically driven by power drawn from a pair of overhead wires via a troller device. His 550 meter line, which ran through the streets of Hallensee in Berlin, can be seen as the forerunner of the trolley bus. Essentially a hybrid of the streetcar and the bus, a trolleybus draws its power from a pair of overhead wires, but then technically works in much the same way as a streetcar, save for the fact that it is not bound to any rails. Therefore it benefits from the power of electric traction, but has the added ability to steer around obstacles that would otherwise block a streetcar's progress. A 6-kilometer trolleybus route served the 1900 Paris Exhibition, and the early 1900s witnessed experiments with trolleybuses in Austria, France, Germany, Italy, and Norway, where a 4-kilometer route opened in Drammen on 10 July 1902. Trolleybuses were introduced in Laurel Canyon in the U.S. in 1910, and in Chicago on 8 October 1914.

Most of the development work on the vehicles happened in the U.K. where, on 20 June 1911 the country's first trolleybus services were introduced in Leeds and Bradford. Up to 1926, trolleybuses developed as trackless streetcars; after 1926 they began to develop as electric buses. This change was largely due to the General Manager of Wolverhampton Corporation Transport—Charles Owen Silvers. Wolverhampton's streetcar system had been built using an unusual surface contact method of current collection for which spares became impossible to obtain by the early 1920s. A decision was taken to install overhead wire power distribution, followed almost immediately by a second decision—to scrap the streetcars. Thus, faced with an almost new power distribution system, trolleybus operation seemed to offer the best alternative to streetcars. The rough riding of the vehicles on the first route to Wednesfield, opened in July 1923, made Silvers examine their design. He experimented by putting trolleybus equipment in the latest design of motorbus body, creating a light, lively, and comfortable vehicle and a showcase transport system. By 1929 Wolverhampton had the largest trolleybus system in the world, and Owen Silvers played host to many delegations from overseas transport operators. Indeed, across the world, a number of trolleybus systems owe their existence to an initial visit to Wolverhampton.

Like motorbuses, trolleybus operation was first seen as an adjunct to that of streetcars, but they later became a direct competitor, enabling, as at Wolverhampton, the life of streetcar power systems to be extended. At the end of the twentieth century, trolleybuses could be found operating in many parts of the world, with the largest number of systems to be found in the former Soviet Union (185); followed by Eastern Europe (58); Western Europe (46); China (25); South and Central America (10); North America and Canada (9); East Asia (8); West Asia (2); and Australia and New Zealand (1).

Underground Railways (Subways, Metros)

Underground railways were a nineteenth-century solution to providing rapid urban transit to and from the heart of busy cities, without further clogging the streets. As with other forms of urban transit, they were transformed in the twentieth century by the application of electric traction to their operation. The world's first underground was the Metropolitan Railway, which was built to connect the center of London with its mainline railway termini. It was operated by steam locomotives and opened on 9 January 1863. Ten further railways, mostly built by separate companies between 1868 and 1907, expanded this line to produce the London Underground system, which has over 275 stations and 407 kilometers of underground tunnels, in some cases stacked three levels deep.

London had a monopoly of subways for almost 30 years, during which time it also opened the world's first electric underground railway—the City and South London—on 4 November 1890. This was also the world's first tube railway, but was only 3.01 meters in diameter and 12 meters below ground. Decidedly cramped, it was nicknamed the "sardine-box railway." The world's second tube system in Chicago demonstrated the best alternative to an underground—the elevated railway. Chicago's "El" was in fact four lines, opened between 6 June 1892 and 31 May 1900, the latter also being the first electric line. Elevated railways had been pioneered in New York by Charles T Harvey, who built an experimental single-track cable-powered elevated railway from Battery Place, at the south end of Manhattan Island, northward up Greenwich Street to Cortlandt Street. Dubbed the "one-legged railroad," the 800-meter single track was carried above the street on a row of single columns. Driven by a stationary engine, the cable was a loop that ran between the rails for propulsion of the cars, and then returned under the street. Harvey's railroad opened for business on 1 July 1868.

Whether under- or overground, the advantages of these urban transit systems are that they occupy dedicated, reserved, and often isolated tracks, which they do not have to share with trains belonging to other operators. Add the acceleration of electric traction, and this gives an urban transit system capable of moving millions of people in and out of cities daily. By the close of the twentieth century there were 124 metro systems worldwide, with 48 in European cities, 42 in Asia, 29 in the U.S., 3 in Africa and 2 in Australia. Their construction has also produced many heroic feats of engineering and outstanding examples of modern architecture. Notable examples are:

- The Liverpool Overhead Railway opened on 4 February 1893.
- Budapest's Metro opened in May 1896 between Vörösmartytér in the center and Széchenyifürdö. The second electric underground railway in the world and the first electric line in Europe.

- The Glasgow Subway opened for traffic on 14 December 1896, but closed the same day following an accident, the service not opening for regular working until 19 January 1897.
- The first line of the Paris Métro opened on 19 July 1900, between Porte de Vincennes and Porte Maillot, making it Europe's fourth oldest metro system, and probably the densest.
- The New York City subway system officially opened on 27 October 1904. It forms 1100 kilometers of New York City's 3200-kilometer public transit system, which is the largest in the world.
- The Moscow metro's first line opened on 15 May 1935 between Sokol'niki and Park Kul'tury, with a branch to Smolenskaya, which reached Kievskaya in April 1937, crossing the Moskva river on a bridge. On a normal weekday it carries 8–9 million passengers, some 3,000,000,000 a year, which is more than the London and New York systems combined.

These transit systems often played a vital role in the development of the cities they served. In the latter half of the twentieth century, major North American cities added subway systems to well established cities such as Washington D.C., San Francisco, and Montreal in an attempt to ease bus and automobile congestion by providing faster alternatives.

Suburban Electric Railways

The first railways usually served local areas by accident, as they passed through en route to larger towns and cities. Typically, suburban railways were built in response to urban growth, a generation or so after the first lines. As with other forms of urban transit, suburban railways were revitalized through electrification. A good example is the lines in the southeast of England serving London. These were electrified progressively from 1 December 1909, with most being converted between 1 April 1925 and 2 July 1939. Electric trains are not only more capable of moving people *en masse* than steam hauled ones, they also allow for a more responsive service. Extra trains can be pressed into service in minutes, as opposed to the 2 hours or so required to steam up a locomotive. Since the 1980s, rail operators in Europe and North America have also increased the capacity of their trains by using double-decked carriages. In Japan, employees help "pack" the subways and move as many people as possible on each train.

See also **Electric Motors; Rail, Diesel and Diesel Electric Locomotives; Rail, Electric Locomotives; Transport**

PAUL COLLINS

Further Reading

Aldridge, J. *British Buses Before 1945*. Ian Allan, Shepperton, 1995.

Clapper, C. *The Golden Age of Tramways*. Routledge & Kegan Paul, London, 1961.

Collins, P. *The Tram Book*. Ian Allan, Shepperton, 1995.

Dunbar, C.S. *Buses, Trolleys & Trams*. Paul Hamlyn, London, 1967. Integrated story of urban transit, along with Charles Clapper's book.

Halliday, S. *Underground to Everywhere*. Sutton Publishing, London, 2001.

Hanscom, W.W. *The Archaeology of the Cable Car*. Socio-Technical Books, Pasadena, CA, 1970.

Howson, H.F. *London's Underground*. Ian Allan, Shepperton, 1986.

Kinder, R.W. Revision of Dendy Marshall's *History of the Southern Railway*. Ian Allan, Shepperton, 1968.

Whipple, F. *Electric Railway*. Orange Empire Railway Museum, Perris, CA, reprint of 1889 publication.

Useful Websites

David Pirmann's encyclopedic site on the New York Subway is one of the oldest and most comprehensive websites, but many others provide information on individual transit undertakings: http://www.nycsubway.org.

V

Valves/Vacuum Tubes

The vacuum tube has its roots in the late nineteenth century when Thomas A. Edison conducted experiments with electric bulbs in 1883. Edison's light bulbs consisted of a conducting filament mounted in a glass bulb. Passing electricity through the filament caused it to heat up and radiate light. A vacuum in the tube prevented the filament from burning up. Edison noted that electric current would flow from the bulb filament to a positively charged metal plate inside the tube. This phenomenon, the one-way flow of current, was called the Edison Effect. Edison himself could not explain the filament's behavior. He felt this effect was interesting but unimportant and patented it as a matter of course. It was only fifteen years later that Joseph John Thomson, a physics professor at the Cavendish Laboratory at the University of Cambridge in the U.K., discovered the electron and understood the significance of what was occurring in the tube. He identified the filament rays as a stream of particles, now called electrons. In a range of papers from 1901 to 1916, O.W. Richardson explained the electron behavior. Today the Edison Effect is known as thermionic emission.

Two further experiments on the Edison Effect were significant contributions to the development of vacuum tubes. John Ambrose Fleming, a British student of James Clark Maxwell, working for the British Wireless Telegraphy Company, attached a bulb with two electrodes to a radio receiving system in order to improve the reception of wireless radio signals. Fleming patented the first electronic rectifier, the diode, or Fleming Valve in 1905. The vacuum tubes were indispensable to switch currents. For example, the Fleming Valve was incorporated in Guglielmo Marconi's wireless system in which communications could take place over great distances. The most serious limitation of the Fleming Valve, however, was that it was insensitive to changes in the concentration of occurrence in electromagnetic radiation. Moreover, the valves consumed large amounts of power.

In 1906 Lee de Forest invented the vacuum tube that became the most important electronic device in the first half of the twentieth century. To Fleming's tube he added a third electrode called a grid—a network of small wires surrounding the cathode to control the stream of electrons. This was the original " triode:" a glass tube containing a filament that heated a plate (the cathode) that emitted electrons, a collector plate (the anode) that collected the electrons, and a metal grid in between. The current flow from the cathode to the anode was highly dependent on the voltage of the grid and the fact that the current drawn by the grid was very low. The name triode came from the three active elements of the device—the anode, the cathode, and the grid. This tube was used as the controllable valve in electronic circuits; for example, to amplify music and voice and to make long-distance calling practical. De Forest called the tube the "Audion," which he patented in 1907 (see Figure 1). Fleming, who was aware of de Forest's activities, had disputed the American's claims to originality, which meant a lifelong enmity between the two inventors.

The first tubes were poorly constructed and able to handle only low levels of power. A further improvement of the vacuum tube came from Irving Langmuir who was working at the Research Laboratory of General Electric. Like some

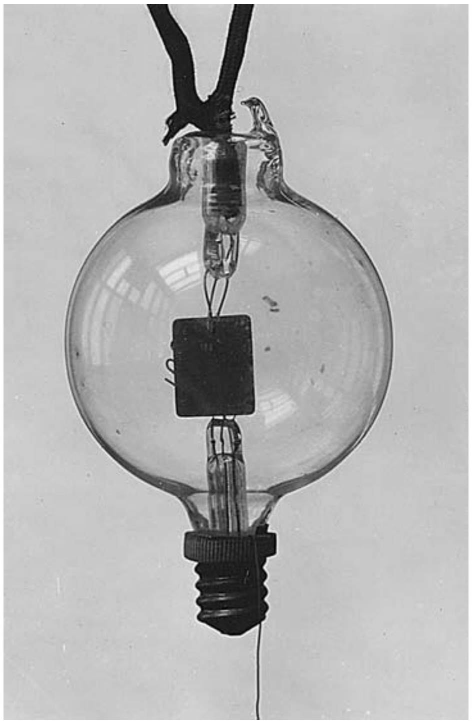


Figure 1. Early DeForest Audion 1906. [*Courtesy of the David Sarnoff Library*].

other physicists at this time, de Forest believed that electron emission in high-vacuum conditions was impossible. Langmuir, however, studied the emission of electrons by hot filaments in a vacuum, because he realized that the Audion did not require the presence of gas to operate. This led to the invention of the high-vacuum electron tube in 1912 that could operate at much greater power levels. These tubes were used as current amplification devices for radio and telephone systems.

In addition to practical problems such as inadequate vacuum pumping, mechanical resonance, and poor welds, the vacuum tube had a major problem: the irregular arrival of electrons at the anode of thermionic tubes. Technologists were trying to solve this problem by adding extra grids to the tube. In 1919 Walter Schottky invented the first multiple grid vacuum tube, the Tetrode. He had found that the fundamental source of noise would be caused by the randomness of the emission from the cathode. Therefore he added a second grid, called the screen, to prevent the tube from producing unwanted oscillations. Bernard D. H. Tellegen, who worked at the Philips Research Laboratory in the Netherlands, made a further improvement to the multiple grid vacuum tube. He invented an advanced receiver tube known as the Pentode. His Pentode tube was a thermionic valve with five electrodes. After Tellegen developed the tube, the most common way of suppressing secondary emission was by using a suppressor grid, placed between a screen grid and the anode. The suppressor grid was maintained at the filament potential level, and it was able to hold back secondary electrons. In 1926 a patent was requested for this invention, and in the same year Pentodes were built into radios.

In 1920 Albert W. Hull of General Electric invented a special vacuum tube called the magnetron. The magnetron is a tube in which the electron beam is controlled by a magnetic field, making extremely high frequencies possible. It's invention led to the development of radar. A major invention was also made by two brothers—Sigurd and Russell Varian—at Stanford University, who applied the principle of velocity modulation to the tube. This high power microwave oscillator used a linear electron beam. It was called the klystron and was built into British and U.S. radar in World War II.

Although many improvements were made, the vacuum tubes still used a lot of electrical power, most of which ended up as heat, thus shortening the life of the tube. This became a big problem in the 1930s because the apparatus for radio, telephony, and computing needed receivers that required high power. Early digital computers like the ENIAC, which was built for the U.S. Army in the 1940s but only completed after the close of the war in late 1945, had about 19,000 vacuum tubes in them, which produced too much heat. The invention of the transistor meant that smaller, reliable, and less power-hungry devices could be produced. Although vacuum tubes are still used, the success of the transistor replaced the vacuum tube technology. After World War II, the transistor became the key to further developments in electronic technology.

See also **Radio Receivers, Early; Radio Receivers, Valve and Transistor Circuits; Transistors**

KEES BOERSMA

Further Reading

Aitken, H.G.J. *The Continuous Wave: Technology and American Radio 1900–1932*. Princeton University Press, Princeton, 1985.

De Forest, L. The Audion—Its Action and Some Recent Applications. J.Franklin Inst., 1920.

Dunsheath, P. *A History of Electrical Engineering*. Faber & Faber, London, 1962.

MacGregor-Morris, J.T. *The Inventor of the Valve: A Biography of Sir Ambrose Fleming*. London, 1954.
Tyne, G.F.J. *The Saga of the Vacuum Tube*. SAMS, Indianapolis, 1977.

Vertical Transportation (Elevators)

Despite the popular concept that the elevator was born at the Crystal Palace Exposition, it actually originated in New York City in 1853 when inventor Elisha Graves Otis first successfully demonstrated his revolutionary new concept—the elevator safety gear or break—which was to allow passengers to travel with safety. The true modern passenger elevator was conceived due to a catastrophic event in 1871 known as the Great Chicago Fire, when a three-day fire razed the city to a desolated wilderness on the plains of Illinois. This fateful day on the 8 October 1871 pinpoints exactly the beginning of the modern elevator.

From the ashes of the old Chicago, the City Fathers made the bold decision that the new City should be constructed from modern materials. The most critical of these forward-looking inventions was the rolled steel joist (RSJ), and by splicing, riveting, and welding together these steel sections, the new construction technique was unveiled. The population explosion in North America required that the new Chicago had to be rebuilt vertically, and with the need to build higher, a new form of elevator was required. The water hydraulic elevator concept had been used since the 1830s, but this design had its limitation on vertical rises due to the restrictions in pressure and the supply of high-pressure water mains. This restricted the vertical rise of the elevator and effectively limited the height of the building. The steam-driven elevator also had its origins in this period, with the first recorded use in 1835.

The first modern passenger elevator in a commercial building was installed in the ten-story Home Insurance Building in Chicago, completed in 1884. Some 13 years after the Great Fire, this building used a steel structure of half cast iron and half steel with supporting masonry walls. The modern elevator age was born with the advent of basic skyscraper technology.

Although the dynamics and control systems of these early elevators were only just evolving, the finishes, materials, craftsmanship, and designs of elevators were in the "Golden Age," when second best was not good enough. With the advent of more complicated control systems and additional safety equipment, components had to be fitted in the elevator shafts and on the elevator cars. This coincided with the beginning of the mass production era, when the large elevator manufacturers realized that this technique enabled more elevators to be built. In addition, new building fabrication techniques also required the elevator engineer to develop new vertical transportation techniques to move passengers higher, faster, and more safely.

Elevators in apartment blocks, banks, and commercial buildings with their imposing entrance lobbies were the symbols of wealth and power. Buildings became active, and they required vertical motion for commerce, bringing spaces alive, and moving people skyward not only rapidly but with grace. The elevator ride was becoming a travel experience to be enjoyed by all. In 1893 the five-story Bradbury Building in Los Angeles brought the elevator into the glass-roofed space spanning the gallery balconies. These elevator cars used open metal designs of the first true panoramic elevator in what John Portman would classify 60 years later as an atrium.

As the new quicker methods of solid elevator shaft construction were being adopted, the passenger's view in and out of the elevator was becoming more and more restricted and the use of glass and decorative materials in elevators was being quickly phased out. For half a century people were moved at ever-increasing speeds, with no views in or out, although the elevator finishes were in some instances very elaborate. The norm, however, was for metal panels (stainless steel, brass, or bronze) or painted finish with the occasional wood panel or mirror. A notable exception to this style was the 1936 Frank Lloyd Wright design for the Johnson Wax headquarters building in Racine, Wisconsin. There, circular observation elevators were used to move travelers through the gallery levels.

It was not until 1962 that architect John Portman had the inspiration to provide elevators in the atrium of a hotel he was designing, not in solid shafts but in an open environment within the building. This revolutionary elevator can be dated exactly to 1965 when the Regency Hyatt, Atlanta opened its doors to the first guests as well as the simply curious.

By the 1980s a new approach to wall-climbing elevators was being adopted. The "flying keyhole" and "birdcage" designs had run their design life, except in hotels, shopping malls, department stores, and speculative commercial developments where corporate and standardized design policies were used. The honest engineering approach was now in vogue with adventurous architects and designers, in which the elevator components and their function were highlighted and expressed.

From the mid-1980s, panoramic elevators had a very high percentage of glass in all visible surfaces to the elevator car. Old skills and craftsmanship were being rediscovered, and computer technology was being incorporated extensively in the control, drive, and design processes to provide not only a good elevator ride but also an experience for the passenger.

Architect Richard Rogers exploited this "total ride" concept in the heart of London in 1986 with his twelve-story neo-Gothic Lloyds of London, where the three elevator towers not only incorporate the elevators but are also fully external. Elevators had to open onto the building lobbies without any of the external environmental conditions penetrating the occupied space of the building. Elevator engineers had to develop new systems for air conditioning, water proofing, and designs to prevent ice and snow build-up on vital safety equipment, while still maintaining reliability, passenger confidence, and safety. This totally new design approach, which celebrated engineering components, had become an essential part of the design of the elevator.

Architects and engineers have always pushed the frontiers of design, none more so than I.M. Pei's Grand Louvre in Paris. The disabled-access elevator within the glass pyramid demonstrates that innovation and excellent design can be achieved on a two-stop low-rise elevator. This elevator is not only a moving sculpture wrapped around a spiral staircase but also, more importantly, a very functional elevator, offering visitors with a disability the opportunity to visit this world famous museum.

The elevator continues to bring new challenges and design opportunities that have been excluded from the elevator engineer's design brief since the advent of the skyscraper, when advanced engineering concepts of their day were required to build structures higher and make elevators travel faster than previously considered safe. Today we understand about height and speed; we now need to know and explore the experience, not in terms of being frightened but in terms of enhancing and experiencing the architecture and design of buildings. The kinesthetic experience (the experience of a body moving through space) will become a vital consideration in future elevator design. Elevator travel has become not just a necessity of the modern city and perhaps even the home but a totally pleasurable experience for the passenger's senses.

See also **Skyscrapers**

ROGER HOWKINS

Further Reading

Elevators. McGraw-Hill, 1927, 1935, and 1960.
Jallings, J.H. *Elevators*. 1918. Republished by Elevator World, 1995.
Strakosch, G. *The Vertical Transportation Handbook*. Wiley, 1983.
Installation Manual. NEMI, 1964.
Howkins, R.E. *Lift Modernisation Design Guide*. Elevator World, 1998.
Philips, R.D. *Electric Lifts*. Pitman, 1973.
Elevator World: U.S.-based monthly magazine.
Elevation: U.K.-based quarterly magazine.

Video Recording, *see* **Television Recording, Tape**

Vitamins

Vitamins are organic compounds essential in minute quantities for the maintenance of health and normal development in most animals and some plants. Many vitamins act as coenzymes or precursors of coenzymes to aid in regulating metabolic processes to produce body tissues or to store or release energy. Others are related to the sterols and hormones. Many vitamins are synthesized by plants and are present in the food intake of animals, though a few are synthesized in the animal body. The existence of such "special factors" in food has long been suspected; for example it has been known since the seventeenth century that fresh fruits and vegetables in the diet prevent scurvy. In 1890 the Dutch physician Christiaan Eijkman showed that beriberi resulted from eating "polished" rice, or rice from which the husks had been removed. When the husks were added to the diet, the disease disappeared, and it was recognized in 1901 that beriberi was due to a dietary deficiency.

Modern vitamin theory dates from 1912 when Sir Frederick Gowland Hopkins showed that animals did not thrive when fed on carefully purified fats, proteins, carbohydrates, mineral salts, and water; but that a very small addition of milk was enough to render the purified diet adequate. In the same year, Casimir Funk at the Lister Institute in London found that the anti-beriberi factor in rice-husks was an amine and propounded his "vitamine" (vital amine) theory, linking vitamin deficiency to other diseases including pellagra and rickets (rachitis). The term was applied to "accessory factors" generally, but as various chemical structures and functions were identified among these compounds and many were found not to be amines at all, they were called vitamins.

Table 1 Recognized vitamins and growth-factors.

Vitamin	Chemical Name	Biological Function	Occurrence
A fat-soluble (discovered 1909) (synthesized 1947)	Retinol	Absence causes night blindness and epithelial tissue becomes keratinous	Egg yolk, milk, fish liver oils
D fat-soluble D2 D3 (disc.1918) (synth. 1959)	Related to steroids ergocalciferol cholecalciferol	Essential for bone and tooth structure; prevents rickets in children and osteomalacia in adults	Egg yolk, milk, fish liver oils. Formed by UV irradiation of sterols
B water-soluble B1 (disc.1897) (synth. 1936) B2 (disc.1933) (synth. 1935) B3 (synth.1941) B6 (disc.1934) (synth. 1939) B12 (disc. 1948) (synth. 1972)	 Thiamine Riboflavin Nicotinic acid (Niacin) Pyridoxine Complex cobalt compound Cyanocobalamin	 Vertebrate nutrition Growth-promoting factor Pellagra preventive Absence causes skin lesions Anti pernicious-anemia factor. Essential for normal blood formation	 Yeast, seed-germs, eggs, liver, flesh and some vegetables Yeast, vegetables, milk, liver Yeast, rice, heart, muscle and liver (nicotinamide) Liver; widely in animal tissues
C water-soluble (disc.1912; synth.1933)	Ascorbic acid	Preventive for scurvy; food antioxidant	Fruits; green leafy vegetables; made synthetically
E fat-soluble (disc.1922; synth.1939)	Tocopherols	Animal nutrition	
H (disc.1931; synth.1943)	Biotin	Protects against toxins in raw egg white	Egg yolk; liver; yeast
K fat-soluble K1 (disc.1929) (synth; 1939) K2	 Phylloquinone Menaquinone	Produce prothrombin essential for blood-clotting	Synthesized by intestinal bacteria in animals
Pantothenic acid (disc. 1933) (synth. 1940)	Pantothenic acid	Important in animal nutrition; present in coenzyme A	Found widely in animal tissues
Folic acid water-sol. (disc.1941; synth.1946)	Folic acid	Prevents anemia	Green leaves; some animal tissues

In 1915, Americans E. V. McCollum and M. Davis showed that there were at least two kinds of vitamins, one soluble in fatty and the other in nonfatty foods. They were later called vitamin A and vitamin B, respectively, and each has since been found to be a group of compounds with related functions. The water-soluble group includes vitamins B and C. The vitamin B complex contains a large collection of compounds, all of which are essential constituents in various enzyme systems. Similarly the fat-soluble vitamin group is now known to include vitamins A, D, E, and K. The physiological functions of most of the vitamins have been examined in detail and their molecular structures have been determined, but their dietary importance has ensured the everyday use of the original alphabetical classification.

Water-soluble vitamins are absorbed in the intestine and carried in the blood to the tissues. They are distinguished from each other by the degree of solubility, a factor that influences their route in the body. In their free state, the B vitamins are inactive, and they must go through several chemical processes before they can perform their functions in the body. When combined with other substances, they are changed into their functional, or coenzyme, form and can then combine with proteins to form active enzymes that catalyze

various metabolic and regulatory processes. The fact that vitamin C prevents and cures scurvy is well known, but the vitamin is also essential for the growth of bones and teeth, for the maintenance of subcutaneous tissues and the walls of blood vessels, and for the healing of wounds. A controversial theory suggests that the intake of large quantities of vitamin C can prevent or cure the common cold, but there is no clear evidence to support this claim. When the intake of water-soluble vitamins exceeds the bodily requirements, they are stored to a limited extent in body tissues, but most of the excess is excreted in the urine.

The fat-soluble vitamins are absorbed with the help of bile salts and are carried through the body in the lymphatic system. The body stores larger quantities of fat-soluble vitamins than of water-soluble ones. The liver provides the chief storage tissue for vitamins A and D; vitamin E is stored in body fat and in the reproductive organs. Vitamins of this group perform various functions. For example, vitamin A combines with proteins in the retina of the eye to aid night vision, though it may have other functions as yet uncertain. The anti-rachitic factor was labeled vitamin D; it is essential to growth, especially in calcium metabolism for bone formation and the avoidance of ricketts. Produced in the skin on exposure to sunlight, vitamin D is one of the few vitamins synthesized by animals. In 1922 a new factor called vitamin E, which facilitates animal growth and maintains fertility in rats and some other species, was identified. Vitamin K is essential for the enzymatic processes of blood clotting.

The number and variety of recognized nutritional factors is continually growing, though not all are essential to animal or human nutrition. Moreover, the definition of a vitamin is imprecise because the metabolic functions of enzymes, coenzymes, sterols, hormones and vitamins are closely related, and much detailed research is required to differentiate between these and other groups of biochemical molecules. Vitamins are not distributed equally throughout nature, nor do they perform the same functions in all species. Sometimes a compound that can act as a vitamin is present in combination with another compound that prevents its absorption and so destroys its activity. Both plants and animals are important natural sources of vitamins for human health, and the more restricted the diet the more likely it is that one or more vitamins will be lacking.

The vitamin requirements of most organisms are fairly well known, but there is not uniform agreement about the requirements for a healthy human diet. Differences arise due to the various ways requirements are determined and to the scanty data available for some of the vitamins. Studies in this field give rise to the subject of human nutrition in which the quantities of the various vitamins figure importantly, along with minerals and other essential trace elements in the diet. It is generally thought that a balanced diet supplies all the vitamins needed for a healthy lifestyle, but this is not necessarily true. Long storage of fruits and vegetables after harvesting may result in the loss of vitamin C due to oxidation. Washed vegetables lose water-soluble vitamins; heating and overcooking destroys them. Manufactured foods therefore often require added vitamins to replace such losses. Sometimes extra vitamins are added to raise the proportion above its natural level, and many people also take vitamin supplements. Milk fortified with vitamin D was first introduced by the Borden Company in the U.S. in 1933. As vitamins were isolated and then synthesized, they could be manufactured by pharmaceutical companies. Vitamin C was first to be synthesized in a laboratory in 1933, followed by vitamin B2 in 1935. By 1936, vitamin and iron supplement sales were widely available in the U.S. In Britain, though not in the U.S., vitamin supplements are classified as drugs.

The chemical structures of all the known vitamins have been determined and most can now be manufactured synthetically by chemical or biochemical processes. The quantity of each product is governed by economic considerations and wide variations in annual production occur. Thus, only about 10 tons of vitamin B_{12} are manufactured each year, but 50,000 tons of vitamin C (ascorbic acid) are made. Manufacturers have developed variations in the synthetic processes, most of which are the subject of trade secrets. Synthetic vitamins produced in bulk are used in the food industry, and large quantities are added to animal feed, but there are also other uses. For example, some vitamins act as antioxidants, and this property has been used in the manufacture of plastics. A small but growing market is also developing for vitamins that benefit the skin, and vitamins are beginning to appear in cosmetics, skin creams, lotions, and shampoos. Research to identify new vitamins and growth factors continues along with efforts to discover more economical methods of manufacture and new uses for the known vitamins.

N.G. Coley

Further Reading

Apple, R.D. *Vitamania: Vitamins In American Culture*. Rutgers University Press, New Brunswick, 1996.

Brewer, S. *The Daily Telegraph Encyclopaedia of Vitamins, Minerals and Herbal Supplements*. Robinson, London, 2002.

Diplock, A.T. *Fat-Soluble Vitamins: Their Biochemistry and Applications*. Heinemann, London, 1985.

Djerassi, D. *The Role of Vitamins in Skin Care*. Hoffmann–La Roche, Nutley, NJ, 1993.

Dyke, S.F. *The Chemistry of the Vitamins*. Wiley/Interscience, London, 1965.

Griffith, H.W. *Vitamins, Herbs, Minerals and Supplements*. Fisher, Tucson, AZ, 1998.

Harris, L.J. The discovery of vitamins, in *The Chemistry of Life: Lectures on the History of Biochemistry*, Needham J., Ed. Cambridge University Press, Cambridge, 1970, pp. 156–170.

Kiple, K.F. and Ornelas, K.C., Eds. *The Cambridge World History of Food*. Cambridge University Press, Cambridge, 2000, vol.1, pp. 741–784.

Marks, J. *A Guide to the Vitamins: Their Role in Health and Disease*. MTP Press, Lancaster, 1975.

Mervyn, L. *Thorson's Complete Guide to Vitamins and Minerals*, 3rd edn. Thorsons, London, 1986.

Reinhard, T. *The Vitamin Source Book*. Lowell House, Los Angeles, 1998.

W

Warfare

Twentieth-century warfare begins with World War I (1914–1918) even though this conflict had more in common with wars of the previous century than it did with those that followed. The Great War opened with maneuvers by huge field armies that culminated in frontal assaults by masses of infantry. After only a few months of mobile warfare, heavy casualties forced opposing armies to take shelter in trench systems that stretched all across France. Faced with the ensuing stalemate of the trenches, both sides adopted an attrition strategy that would defeat the opposition by bleeding his manpower and depleting his material resources. The strategy finally succeeded when an exhausted Germany surrendered in November 1918.

In spite of its similarity to nineteenth-century warfare, World War I witnessed several new developments, most notably the airplane, the tank, and the truck. Between 1919 and 1939, the implications of these new developments were worked out, producing new operational approaches that transformed warfare.

During World War II (1939–1945), European land warfare was dominated by mobile armored forces that swept back and forth across the continent. While armies fought on the ground, air forces contended for control of European skies. In this massive air war, Allied bombers devastated Germany's industrial base and population centers.

Meanwhile, in the Pacific region, the war centered on aircraft carrier task forces that battled each other and supported amphibious operations. The war started with Japan conquering much of the western Pacific, only to be pushed back by superior Allied arms and forced to surrender when American B-29 bombers dropped the only two atomic bombs ever used in war.

By the time World War II ended in the Pacific, Japan's military resources had been severely reduced by Allied military actions. The reduction of Japanese resources, along with the progressive weakening of Germany in the European theater, suggest that World War II, like the first, was an attrition war in which industrial capacity was as important as military forces.

Two of the most revolutionary developments of World War II were the atomic bomb and the long-range ballistic missile. When more fully developed and mated to each other during the Cold War (1946–1991), they became what is perhaps the most revolutionary weapon in military history, the nuclear-tipped, intercontinental-range ballistic missile (ICBM). In the end, the Cold War was another attrition conflict, ending with the economic exhaustion and collapse of the Soviet Empire.

The end of the Cold War reduced the tensions that had kept nuclear strike forces on hair-trigger alert since the 1950s. Although nuclear weapons still existed, relations between the U.S. and the Russian Federation that emerged from the defunct Soviet Union were no longer based on mutually assured destruction (MAD), as both sides reduced their nuclear forces and the U.S. continued developing missile defenses.

While there were a number of significant "limited" wars during the twentieth century, the five major episodes described above unleashed the greatest national energies. These energies were molded into major new military systems through the process of command technology that is rooted in England of the 1880s according to historian William McNeill. Before this time, weapons were

either developed in government-owned arsenals or by private entrepreneur inventors. A major change began in 1886 when the British Admiralty, dissatisfied with the performance of the government arsenal at Woolwich, started contracting with private arms makers for the development of new weapons. Under this approach, the Admiralty established the specifications for a new weapon and effectively challenged the contractor to produce it. This contracting system marks the beginning of command technology. Tantamount to invention on demand, this process of state-sponsored research and development spread throughout the West, becoming the dominant paradigm for weapons acquisition by 1945.

One product of command technology during World War I was the tank, which was developed by the British to cross fire-swept terrain between the trenches and breech the German defenses. While the tank proved capable of completing its mission, its successes were limited due to technical limitations and a lack of understanding of how best to use the new weapon.

The principal enabling technology for the tank was the internal combustion engine, which also powered World War I trucks and airplanes. The former improved logistics by connecting troops in forward positions with railheads and supply depots in the rear. The latter opened an entirely new realm of warfare and, over the course of the war, suggested all the missions the airplane would perform in future wars.

Building on the lessons of World War I, air power advocates used the period between the two world wars to develop a rigorous body of air power doctrine. At the same time, the world's leading powers developed aircraft of increasing capabilities to execute the missions defined in their doctrines.

The U.S. emphasized long-range bombers to execute daylight, precision bombardment—the dominant doctrine in America's air force. England also developed bombers, but she also pursued fighter development because of the threat posed by the air force of a rearming Germany. In addition to bombers, Germany developed tactical aircraft to support its new approach to ground warfare—Blitzkrieg.

The basic ideas behind Blitzkrieg had emerged by the end of World War I, as the capabilities of tanks and aircraft improved. After the war, the Germans developed these ideas further and mated them to the panzer division, which included tanks, mechanized artillery, and motorized infantry. Through radio communications, these elements were integrated into coherent units that also used their radios to coordinate supporting air attacks by Germany's tactical fighters. Using their air support, the panzers would execute deep, penetrating attacks to unbalance opponents and keep them from shoring up their defenses once these had been breeched.

World War II in Europe opened with Blitzkrieg attacks that swiftly overran Poland in 1939 and France in 1940. It ended with Allied air forces supporting mechanized operations that pushed German forces out of their conquered territories prior to overrunning Germany itself.

While the Germans were perfecting Blitzkrieg, naval officers around the world were integrating aviation into naval operations. This entailed developing true aircraft carriers with landing decks that ran the full length of vessels, allowing aircraft to both take-off and land on the carrier. The advent of these carriers prompted a debate over which ship, the carrier or the battleship, would dominate the next war.

This question was settled decisively at Pearl Harbor on December 7, 1941, when Japanese carrier aircraft damaged or sank every American battleship in the harbor. Throughout the remainder of the war in the Pacific, the principal measure of naval power was the carrier task force in which battleships, cruisers, and destroyers used their firepower principally to protect their carriers from attack by enemy planes and submarines. The impact of the carrier on naval warfare is clearly illustrated by the May 1942 Battle of the Coral Sea, the first naval engagement in which the surface forces never sighted each other. Throughout the remainder of the century, the carrier task force dominated naval operations.

Three years before the Battle of the Coral Sea, physicist Albert Einstein alerted U.S. President Franklin Roosevelt to the potential of nuclear fission. After having this concept evaluated by a panel of scientists, Roosevelt launched the Manhattan District Project to develop an atomic bomb.

There were two major facets to this project: developing an industrial base to produce fissionable materials and designing the bomb itself. On July 16, 1945, less than three years after the project began, the world's first atomic bomb was detonated in New Mexico. Within a month, the U.S. had dropped two atomic bombs on Japan, forcing the Japanese to surrender. About a decade before the U.S. launched its atomic bomb project, the Germans began work on what would become the world's first long-range ballistic missile. In 1937, this program was greatly expanded with the

establishment of a vast, new rocket development center at Peenemunde. The German program employed several hundred scientists and technicians who were supported by a large budget that could be coupled to Germany's industrial base and its university research facilities through a flexible contracting system.

This rocket program is a classic example of command technology. Guided by specifications established by the army's ordnance office, the Peenemunde team made rapid progress after 1937. In June 1942, the team completed the first successful test of the V-2 rocket, which became the world's first operational long-range missile when it began hitting allied cities in September 1944. This choice of targets, which was dictated by the missile's inaccuracy and the limited size of its warhead, meant that the V-2 was essentially a terror weapon with little real military value.

Immediately after World War II, both the U.S. and the Soviet Union absorbed German rocket developments and began working energetically to produce long-range missiles that could be used for military purposes. A major breakthrough came in the 1950s when both countries demonstrated the ability to produce thermonuclear bombs. This meant that warheads could be made that were light enough to be carried by a missile, yet powerful enough to compensate for missile inaccuracies. Moreover, the advent of the hydrogen bomb ushered in an era of "nuclear plenty," since fusion fuel is plentiful and inexpensive when compared to fission fuel.

At the same time, work was progressing on inertial guidance systems that would be much more accurate than the system used in the V-2. A major breakthrough here was the development of more sensitive inertial measuring units that were based on complex mechanical structures, computer advances, and improved electro-optical technologies.

The simultaneous resolution of guidance and warhead problems made the ICBM feasible. Paradoxically, because these weapons could destroy civilization, the doctrine governing their employment, mutual assured destruction (MAD), aimed to deter their use. MAD required each side to have enough nuclear weapons to absorb a nuclear attack and still be able to inflict unacceptable losses on the attacker.

America's first ICBM became operational in 1959. In developing this missile, the U.S. Air Force pioneered a new management discipline that was based on insights into the functioning of complex weapons.

Until well into the nineteenth century, weapons were largely simple, stand-alone devices. However, by World War I, they were often amalgams of complex components as in the case of the giant dreadnought class battleships that dominated naval warfare during the first two decades of the twentieth century.

During the World War II, air defenses and aircraft carriers raised the complexity of weaponry another order of magnitude. It was at this point that the pioneers of operational analysis made the point that optimizing the performance of complex weapons required a thorough understanding of how their components interacted with each other and with their operational environment. Assuring a proper "fit" between system components became the work of systems engineering. Bringing operational analysis and systems engineering together to create an effective weapon was the function of systems management, a discipline that was more fully developed and formalized in the U.S.'s huge ICBM program that was launched in the 1950s. The success of the ICBM program transformed systems management into the principal paradigm for managing major weapons programs, including those for self-guided and precision-guided munitions (PGMs).

A major inspiration for self-guided munitions was the airplane. Before the advent of artificial sensors, computers, and advanced servo motors, the presence of a pilot offered one means, beyond initial aiming, to guide a weapon to its target. Indeed, one of the best known early efforts to achieve precision guidance was the Japanese use of suicide pilots who attempted to fly their planes into U.S. ships during the World War II. Less well known are U.S. and German efforts to develop unmanned glide bombs and vertical bombs that could be controlled from the aircraft that dropped them.

Germany's desperate efforts to down Allied bombers near the end of the World War II spawned several innovative concepts in the area of precision-guided surface-to-air missiles or SAMs. Included here was the use of a simple infrared sensor to allow SAMs to home in on hot bomber engines. Another SAM was to have been guided by commands from the ground that reached the interceptor via a thin wire that played out as the missile flew toward its target. Fortunately for Allied bombers, these ideas came too late in the war to be implemented.

More fully developed after World War II, wire-guided missiles were used extensively in limited and regional wars such as the Vietnam War (1965–

1973) in which an American wire-guided missile achieved an 80 percent hit rate. Soviet wire-guided missiles were used extensively by the Egyptians to inflict heavy loses on Israeli armor during the early phase of the 1973 Egyptian–Israeli War.

Infrared heat-seeking technology was widely applied in missile guidance after 1945. By 1953, the U.S. had developed the world's first heat-seeking air-to-air missile. Widely used throughout the rest of the century, these missiles generally employed a small, nose-mounted infrared sensor to guide them to the hot engine tailpipe of enemy aircraft. Shoulder-held heat-seekers were also developed to protect soldiers against air attacks. By the end of the century, the spread of these small, portable missiles was causing concern that terrorists might use them against commercial jetliners.

Other precision-guided missiles used radar in their guidance systems. While some were designed for air-to-air combat, others were built to home in on the signal from air defense radars. Radar-guided SAMs also became central to effective air defenses.

Systematic efforts to develop defenses against aircraft began during World War I when the British tried to stop German bomber attacks on England. Twenty years later, with England facing the prospect of air attacks from a rearming Nazi Germany, Sir Robert Watson-Watt advised the British government that reflected radio waves could be used to locate attacking aircraft. This principal became the basis for a radar system that the British began deploying in the mid-1930s. By the time German planes attacked London in 1940, England had deployed an air defense system that used radar plots and radio communications to guide defensive fighters to attacking German planes.

The use of radar here is an important departure. The increasing speed and range of the airplane collapsed time and threatened to deprive the defender of adequate response time. Using instruments such as binoculars and listening devices to increase the power of human senses was no longer adequate for locating an attacking force. Radar marks the first effort in military affairs to extend human perception by using phenomena outside the normal range of man's five senses. The British Chain Home radar system could detect aircraft approaching at an altitude of 6000 meters at a range of 145 kilometers, providing a warning time of 15 minutes for planes flying at 580 kilometers per hour.

Faced with the threat of nuclear-armed Soviet bombers in the 1950, the US. developed a continental-wide air defense system with a forward-based radar system to provide the earliest possible warning of attack. Radar data were fed to computerized control centers that automated the manual process of vectoring interceptors to their targets. These centers could simultaneously track 200 attacking bombers while vectoring 200 interceptors to their intercept points.

As this system was becoming operational, both the U.S. and the Soviet Union began deploying ICBMs. Against these weapons, bomber defenses were essentially useless. The ICBM's speed allowed it to traverse thousands of kilometers in a matter of minutes, further compressing the time for defensive actions. Some way had to be found to recapture the lost response time.

Improved ground-based radars provided fifteen minutes of warning time in the case of an ICBM attack. An additional fifteen minutes were gained by deploying satellite-based, infrared sensors that surveilled enemy missile fields around the clock. More time was recovered by parsing time into picoseconds, providing billions of time units that could be effectively managed by high-speed computers to optimize defensive reactions.

Later missile defense concepts pursued under the Strategic Defense Initiative, which was launched in 1984 by the US., sought to improve the odds for a successful defense by placing interceptor missiles in space. Furthermore, the U.S. pursued various concepts for directed energy weapons, which promised a near instantaneous kill, since beam velocities approached the speed of light. Combining orbiting lasers with space-based interceptors would produce a defense capable of destroying enemy missiles during the boost phase before they released their multiple warheads and decoys.

As the twentieth century was ending, the U.S. was developing an airborne laser that could also destroy ballistic missiles during their boost phases. This weapon also promised to be effective against attacking aircraft.

The high-speed computer, so crucial to the prospects of missile defense, was also central to the development and proliferation of command and control systems after 1950. These systems formed an integrated "picture" of current situations based on information from a wide variety of sources. Included among these sources are battlefield sensors, overhead satellites, electronic intelligence, and units engaged in combat. This picture provided the basis for extending and tightening the control exerted of senior political and military leaders. Computerized systems also played a major

role in managing military logistics, so essential to modern military forces.

Developments such as high-speed computers, lasers, radar, and infrared sensors point toward a fundamentally new departure in twentieth-century weaponry: the creation of advanced military capabilities based on esoteric scientific principles. These principles are generated through abstract, mathematical reasoning and are not readily discoverable through the traditional methods of careful observation and the manipulation of materials. Without the highly mathematical electromagnetic field theory of James Clerk Maxwell there would be no radio or radar. Without the work of scientists like J. J. Thompson, Ernest Rutherford, and Niels Bohr there would have been no atomic theory and no basis for conceiving nuclear fission.

Introducing scientists into the mix of engineers, technicians, and managers that was central to earlier forms of command technology greatly increased government's power to invent on command. As the century was ending, this enhanced form of command technology had created in military affairs a situation similar to what historian Walter McDougall described as a perpetual technological revolution.

Change had become one of the few constants in military affairs. Making effective "transformations" in force structures and doctrines to ensure success in future wars was more clearly than ever a core concern for military professionals and their civilian leaders.

See also **Battleships; Fission and Fusion Bombs; Military versus Civil Technologies; Tanks; Missiles; Radar; Sonar; Submarines, Military; Warfare, Biological; Warfare, Chemical; Warplanes**

DONALD R. BAUCOM

Further Reading

Brodie, B. and Fawn, W. *From Crossbow to H-Bomb: The Evolution of the Weapons and Tactics of Warfare,* revised edn. Indiana University Press, Bloomington, 1973.

Dupuy, T.N. *The Evolution of Weapons and Warfare*. The Bobbs-Merrill Company, New York, 1980.

Huges, T.P. *Rescuing Prometheus*. Pantheon, New York, 1998.

Johnson, S.B. *The United States Air Force and the Culture of Innovation: 1945–1965*. Air Force History and Museums Program, Washington D.C., 2002.

McDougall, W.A. *The Heavens and the Earth: A Political History of the Space Age*. Basic Books, New York, 1985.

McNeill, W.H. *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000*. University of Chicago Press, Chicago, 1982.

Van Creveld, Martin. *Technology and War: From 2000 B.C. to the Present*. The Free Press, New York, 1989.

Warfare, Biological

In addition to the military use of natural or synthesized plant and animal toxins as poisons, biological warfare involves the use of disease-causing bacteria, viruses, rickettsia, or fungi to cause incapacitation or death in man, animals, or plants. Over the course of the twentieth century, biological weapons scientists, engineers, and physicians in various countries adopted existing technological and scientific practices, techniques, and instrumentation found in academic and industrial research to create a new weapon of mass destruction. Unlike the production of nuclear weapons, biological weapons research involves a synergistic relationship between the separate offensive and defensive components of each individual weapon system. Offensive research involves the identification, isolation, modification, and mass production of various pathogenic organisms and the creation of organismal delivery and storage systems. Offensive research is dependent in many cases upon the simultaneous success of a parallel defensive research program involving the creation of vaccines and protective health measures for researchers, military personnel, and civilians. In addition, defensive research involves the construction of accurate detection devices to indicate the existence of biological weapons whose presence can be masked during the initial phases of a natural epidemic.

Combining practices from medical bacteriology and public health such as pure culture technique and vaccination with standard industrial microbiological practices used in the mass production and storage of pharmaceuticals, foodstuffs, and chemicals, military researchers created industrially modified organisms that could be used against specific agricultural, civilian, and military targets. The dual-use nature of much of the technology used to create and store biological weapons, coupled with the availability of this technology on open world markets, has created difficulties for those nation states and international regulating bodies attempting to prohibit the further expansion of this military technology to other states and possibly even to terrorist organizations. The tactical and strategic utility of biological warfare has evolved as a function of the capabilities of specific biological weapons, and it was not until the perfection of airborne delivery systems using aerosolized dry matter by the U.S. and the Soviet Union during the 1960s that biological weapons revealed a destructive potential equal to the strategic threat of nuclear weapons.

Biological weapons were initially conceived for use against agricultural targets as a form of economic and psychological warfare. The creation of directed research programs in industrialized nations beginning during the 1930s indicated the possibility of using biological weapons as antipersonnel weapons and this research has continued to the present day. During World War I, against a backdrop of chemical weapons use by belligerents on both sides, the General Staff of the German Army carried out a secret campaign of sabotage involving injections of anthrax (*Bacilus anthracis*) and glanders (*Pseudomonus mallei*) into military livestock of neutral nations such as the U.S., Romania, Norway, and Spain that resulted in the deaths of numerous horses between 1915 and 1918. Following the war, the 1925 Geneva Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare prohibited "the use of bacteriological methods of war." Unfortunately, many nations interpreted the treaty as reserving the right for retaliation in kind and top secret research was carried out in France, Germany, Great Britain, Japan, Poland, Germany and the Soviet Union during the interwar period.

The onset of World War II accelerated many of these research and development programs and initiated new programs in other countries, most significantly in the U.S. Although German scientists mass-produced nerve gases such as tabun, sarin, and soman, Nazi interest in biological warfare was constrained by Adolf Hitler's distaste for the weapon. Surprisingly little practical research was conducted and Germany was ill prepared for either carrying out or defending against biological warfare. While historical inquiry into the Soviet program during this and later periods continues, it now seems possible that Soviet military forces may have successfully used tularemia (*Francisella tularensis*) as a tactical biological weapon to thwart German advances during the pivotal battle of Stalingrad in 1942. Large-scale research was carried out by Japanese military scientists and physicians in Unit 731 under General Ishii Shiro in Japanese-occupied Manchuria from 1936–1945. Ishii tested the feasibility of anthrax, cholera (*Vibrio cholerae*), dysentery (*Shigella dysenteriae*) and plague (*Yersinia pestis)* through studies involving human experimentation and torture that resulted in the deaths of thousands of political prisoners, captured POWs, and civilians, including woman and children. Offered asylum by American officials after the war, Ishii's research results were added to the world's most advanced biological weapons research program. In cooperation with Great Britain and Canada, the U.S. program encompassed more than 23 different research projects involving specialized laboratory, pilot plant, and manufacturing equipment at top secret military facilities, and through research contracted to university, governmental, and industrial laboratories at a cost of $60 million. Significant research was conducted on a host of biological warfare agents including anthrax, brucellosis (*Brucella melitensis)*, dysentery, glanders, plague, tularemia and various chemical plant growth inhibitors. In addition projects were carried out on vaccination, mass production and storage, the aerosolization of pathogens, and the dynamics of airborne infection.

The U.S. and the Soviet Union continued research on biological warfare throughout the Cold War and during the 1950s and 1960s successfully "weaponized" anthrax, brucellosis, and tularemia as biological aerosols. During this period researchers moved beyond the liquid slurry form of these agents, and combining technologies of drying, freezing, and milling produced agents which could be stored for long periods and be disseminated from various delivery systems including aircraft and missiles to inflict mass casualties. Large-scale ecological testing in the Pacific Ocean involving laboratory animals verified the destructive capability of these weapons over thousands of square kilometers under specific meteorological and ultraviolet conditions. Although a signatory (like the U.S.) to the 1972 Convention on the Prohibition of the Development, Production, and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction, the Soviet Union greatly expanded it's offensive biological weapons program under the Ministry of Defense during the 1970s and 1980s. Under Biopreparat, Soviet officials created the world's largest and most sophisticated research program involving more than 15,000 personnel, six research laboratories, and five large-scale production facilities that incorporated recent discoveries in recombinant DNA technology to create a new generation of biological warfare agents.

An anthrax outbreak in April 1979, in Sverdlovsk (now Ekaterinburg, Russia) resulted from the explosion of one such military production facility and raised questions about Soviet treaty compliance. The breakup of the Soviet Union since 1991 has greatly destabilized the security of these weapons facilities, and questions remain about the movement of former Soviet scientists and materials. In the U.S. during the fall of 2001 the

mysterious use of letters containing domestic weapons grade anthrax contaminated citizens, postal workers, government officials, and members of the news media, resulting in five deaths. Although the difficulty and expense involved in standardizing airborne delivery systems has limited the use of biological weapons as a weapon of mass destruction, the inhalation fatalities resulting from the handling of anthrax contaminated mail indicates that biological warfare and biological terrorism pose a serious threat in the near future.

See also **Warfare, Chemical**

GERARD J. FITZGERALD

Further Reading

Carter, G. and Balmer, B. Chemical and Biological Warfare and Defense, 1945–1990, in *Cold War Hot Science: Applied Research in Britain's Defense Laboratories 1945–1990,* Bud, R. and Gummett, P, Eds. Harwood, Amsterdam, 1999.

Carter, G. and Pearson, G.S. North Atlantic chemical and biological research and collaboration: 1916–1995. *J. Strat. Stud.,* 19, 74–103, 1996.

Geissler, E. and van Courtland Moon, J. E., Eds. *Biological and Toxin Weapons: Research, Development and the Use from the Middle Ages to 1945*. SIPRI Chemical & Biological Warfare Studies, 18, Oxford University Press, Oxford, 1999.

Harris, S.H. *Factories of Death: Japanese Biological Warfare, 1932–45 and the American Cover Up*. Routledge, New York, 1994.

Lederberg, J., Ed. *Biological Weapons Limiting the Threat*. MIT Press, Cambridge, MA, 1999.

Wright, S., Ed. *Preventing a Biological Arms Race*. MIT Press, Cambridge, 1990.

Warfare, Chemical

Popular fiction forecast the use of poison gas in warfare from the 1890s. While an effort was made to ban the wartime use of gas at The Hague International Peace Conference in 1899, military strategists and tacticians dismissed chemical weapons as a fanciful notion. The stalemate of World War I changed this mindset. Under Fritz Haber, a chemist at the Kaiser Wilhelm Institute, Germany's chemical industry began making gas weapons. Compressed chlorine gas in 5730 cylinders was released against French Algerian and Canadian troops at Ypres, Belgium, on April 22, 1915. The gas attack resulted in approximately 3000 casualties, including some 800 deaths. Within months the British and French developed both gas agents of their own and protective gear, ensuring that chemical warfare would become a regular feature of the war.

A variety of lethal and nonlethal chemical agents were developed in World War I. Lethal agents included the asphyxiating gases such as chlorine, phosgene, and diphosgene that drown their victims in mucous, choking off the supply of oxygen from the lungs. A second type were blood gases like hydrogen cyanide, which block the body's ability to absorb oxygen from red corpuscles. Incapacitating gases included lachrymatorics (tear gases) and vesicants (blistering gases). The most notorious of these is mustard gas (*Bis*-[2-chloroethyl] sulphide), a blistering agent that produces horrible burns on the exposed skin and destroys mucous tissue and also persists on the soil for as long as 48 hours after its initial dispersion.

Ultimately chemical warfare had little decisive effect on the outcome of World War I. On the western front, approximately 500,000 casualties resulted from gas, of which less than 5 percent were fatalities. Defensive measures, including new gas helmets and masks, and effective decontamination procedures, kept pace with the appearance of new gases, reducing their long-term effectiveness. By 1917 gas shells were relegated to the status of auxiliary weapons, used to shut down enemy artillery and machine gun positions and disrupt communications.

After the war chemical weapons acquired a sinister reputation far out of proportion to their military effectiveness at the time. Much of this was due to the postwar writings of veterans, who projected their own experiences and fears onto a wider strategic venue. Military strategists like the Italian apostle of air power Guilio Douhet and Britain's Sir Basil Liddell Hart increased the negative perception of gas warfare by describing its use as a strategic terror weapon for use against civilian populations. The offensive use of chemical agents in wartime was outlawed, first by the 1922 Washington Disarmament Treaty, and then more broadly by the 1925 Geneva Protocols on gas and bacteriological warfare.

Despite the international agreements, most countries continued developing chemical weapons in the 1920s and 1930s, led by Germany, the U.S., and Great Britain. American and British scientists worked on improving existing agents, as well as creating new lethal agents using organic compounds such as ricin (the toxic distillate of the castor bean), BTX (botulinal toxin), and CN (chloroacetophenone), a lachrymatory agent. By 1936 German chemists, examining fluorine and phosphorus compounds for the insecticide industry, developed tabun (ethyl NN-dimethylpho-

sphoramidocyanidate), the first nerve gas. This was followed in 1939 by sarin (isopropyl methylphosphonofluoridate). The German nerve gases, identified as G-agents after the end of World War II, were considerably more effective than the asphyxiating gases. Lethal in very small doses, the G-agents were absorbed through the skin and were odorless and colorless. Despite the work during the interwar years, the major combatants refrained from using chemical weapons against each other in World War II, fearing that battlefield use would soon translate into deployment against civilian populations.

After World War II, captured data and samples of the G-agents became the foundation for chemical warfare development programs by American and Soviet researchers. Sarin and VX, a second-generation reformulation of the G-agents, formed the core of the American chemical weapons arsenal until 1969. Following a series of embarrassing incidents, including the accidental release of gas at Dugway Proving Grounds in Utah in 1968, President Richard Nixon formally renounced American production, storage, and use of chemical weapons. The unilateral ban remained in effect until the Reagan administration in the 1980s. Facing a real and perceived Soviet advantage in conventional and strategic weapons, the U.S. resumed production of chemical weapons agents and delivery systems, which included a binary shell system that separated the chemical ordnance into two harmless components for storage, mixing upon explosion into the desired agent. Although the Soviet Union was recognized as maintaining the world's largest stockpile of chemical weapons, many details of the program remained hidden until after 1991, when the scope and extent of Soviet research into their own G-agent derivatives and biological toxin agents emerged. Since the collapse of the Soviet Union in 1991, however, the U.S. and Russia have moved toward chemical weapons disarmament and disposal.

Since World War I, chemical weapons have been used only sparingly, and then generally against weaker opponents without their own offensive or defensive capabilities and against helpless civilian populations. Fascist Italy acknowledged using chemical agents (later found to be mustard gas) during their 1935 invasion of Ethiopia. The Imperial Japanese Army regularly used a variety of agents, including phosgene, hydrogen cyanide, and mustard gas, against military forces and civilians in China from 1937 through 1945. A commercial compound of hydrogen cyanide known as Zyklon-B was also used to grim effect against Jews, Soviet prisoners, gypsies, and other victims in the Nazi death camps.

After World War II allegations about the use of chemical weapons by government or colonialist forces were often reported but rarely substantiated. Beginning in 1962 the U.S. used chemical defoliant agents to eliminate tactical cover for National Liberation Front and North Vietnamese Army forces operating in South Vietnam, as well as to destroy rice fields supporting them in Project Ranch Hand. Between 1962 and 1969, the United States Air Force sprayed 22,336 square kilometers with herbicides, the most common of which was Agent Orange (*n*-butyl 2,4-dichlorophenoxyacetate and *n*-butyl, 2,4,5-trichlorophenoxyacetate). While not deployed as an antipersonnel agent, since the end of the Vietnam War, Agent Orange has been linked to high rates of cancer, birth defects, and sterility among those exposed to it and their descendents. Refugees from Laos and Afghanistan throughout the 1980s described a "yellow rain" that brought illness and death to those exposed to it. The nature of the substance was the subject of debate for some years. One theory maintained the agent was really feces from a large swarm of honeybees, while Western defense analysts pointed to material evidence and eyewitness testimony that indicated the substance was a Soviet-produced chemical agent, T-2-trichothecene mycotoxin.

Iraq employed mustard gas, tabun, and other agents against Iran in the war from 1979 to 1988. Iran responded with its own gas weapons by 1986. Sarin and mustard gas were also used against Kurdish villages in northern Iraq during and after the Iran–Iraq War. The Iraqi chemical warfare program continued to develop agents and delivery systems, including warheads for *al-Hussein* intermediate ballistic missiles during and after the 1991 Persian Gulf War, despite international sanctions and the objections of the U.S. and its allies.

The Iraqi example indicates the trend of chemical weapons as strategic weapons "on the cheap" for developing nations. Relatively inexpensive to produce and adaptable to a variety of delivery systems, chemical weapons can provide a qualitative advantage previously unavailable. The proliferation of skilled technicians and scientists from the former Soviet Union after 1991 exacerbated this problem. By the turn of the twenty-first century, non-state actors in the form of terrorist organizations had arisen. In 1995 the Aum Shunrikyo religious cult released sarin on board commuter trains in Tokyo, killing seven and injuring another 144 people. This act highlighted the ease with which individuals and small groups with technical knowl-

edge and access to chemicals could effectively "home-brew" sophisticated agents.

See also **Warfare, Biological**

BOB WINTERMUTE

Further Reading

Croddy, E., Perez-Armendariz, C. and Hart, J. *Chemical and Biological Warfare: A Comprehensive Guide for the Concerned Citizen*. Copernicus, New York, 2002.

Haber, L.F. *The Poisonous Cloud: Chemical Warfare in the First World War*. Clarendon Press, Oxford, 1986.

Spiers, E.M. *Chemical Warfare*. Macmillan, London, 1986.

Stockholm International Peace Research Institute (SIPRI). *The Problem of Chemical and Biological Warfare,* 5 vols. Almqvist & Wiksell, Stockholm. Humanities Press, New York, 1971.

Useful Websites

Fedorov, L.A. *Chemical Weapons in Russia: History, Ecology, and Politics* [*Khimicheskoye Oruzhiye V Rossii: Istoriya, Ekologiya, Politika*]. Federation of American Scientists, Washington D.C., July 27, 1994: http://www.fas.org/nuke/guide/russia/cbw/jptac008_l94001.htm

Warfare, High-Explosive Shells and Bombs

Among the most baleful of twentieth century technological accomplishments was the vast elaboration of the means for inflicting death and destruction in war. While nuclear and chemical weapons occasioned more revulsion, conventional high-explosive weapons wrought far wider harm.

A revolution began in the nineteenth century with the introduction of rifled cannon and effective explosive shells. This, in turn, brought an escalating contest between weapons and protection both for fortifications and ships. At the beginning of the twentieth century, shells were beginning to move from black powder fill to modern high explosives such as ammonium picrate and trinitrotoluene (TNT). High-explosive (HE) shells needed steel walls thick enough to withstand the shock of firing, limiting weights of bursting charges to no more than about 25 percent of the whole. Depending on the target, they might use either point-detonating or time fuses. The early time fuses continued, as they had in the nineteenth century, to depend on the time taken for a powder train of precut length to burn to its end.

Shells intended for penetration had very hard, rather blunt steel points to break the hard face of the armor, backed by a much softer and tougher body to hold the projectile together through the course of the penetration. A slightly softer cap over the point of the shell might serve to preload the armor's hardened face. Bursting charges generally were no more than 5 percent of projectile weight and were detonated by base fuses intended to delay until penetration had been achieved. Also widely

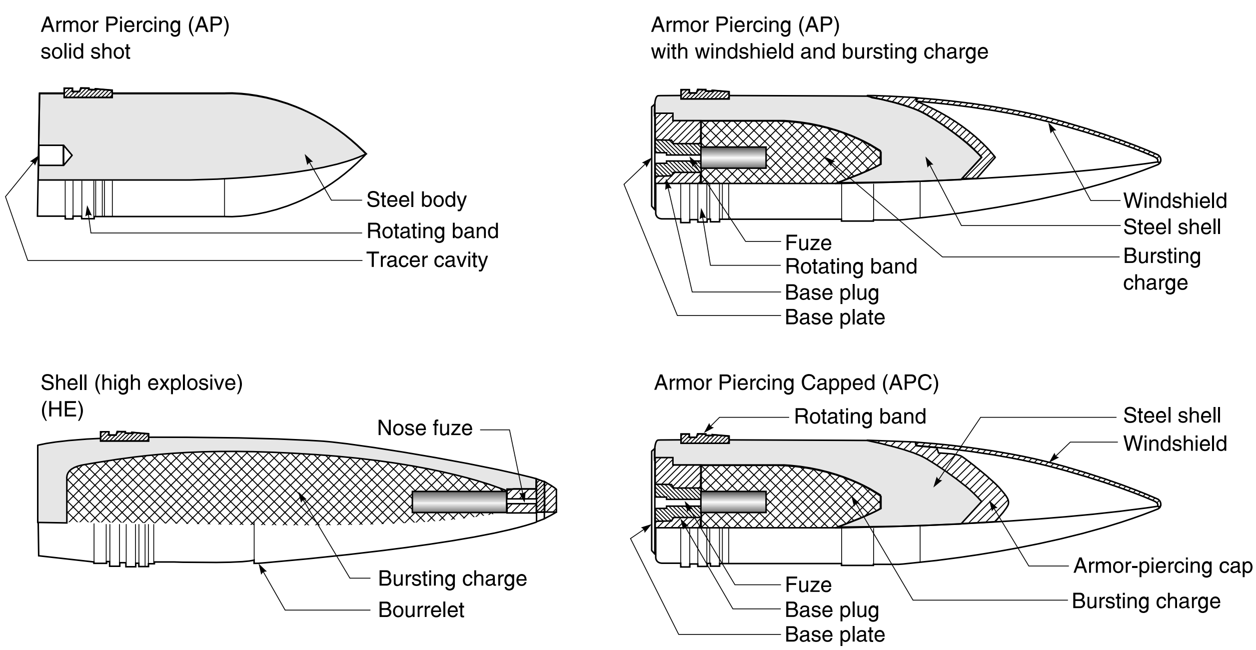


Figure 1. U.S. Army 155 mm World War II projectiles are broadly typical of types used by all nations throughout the century.

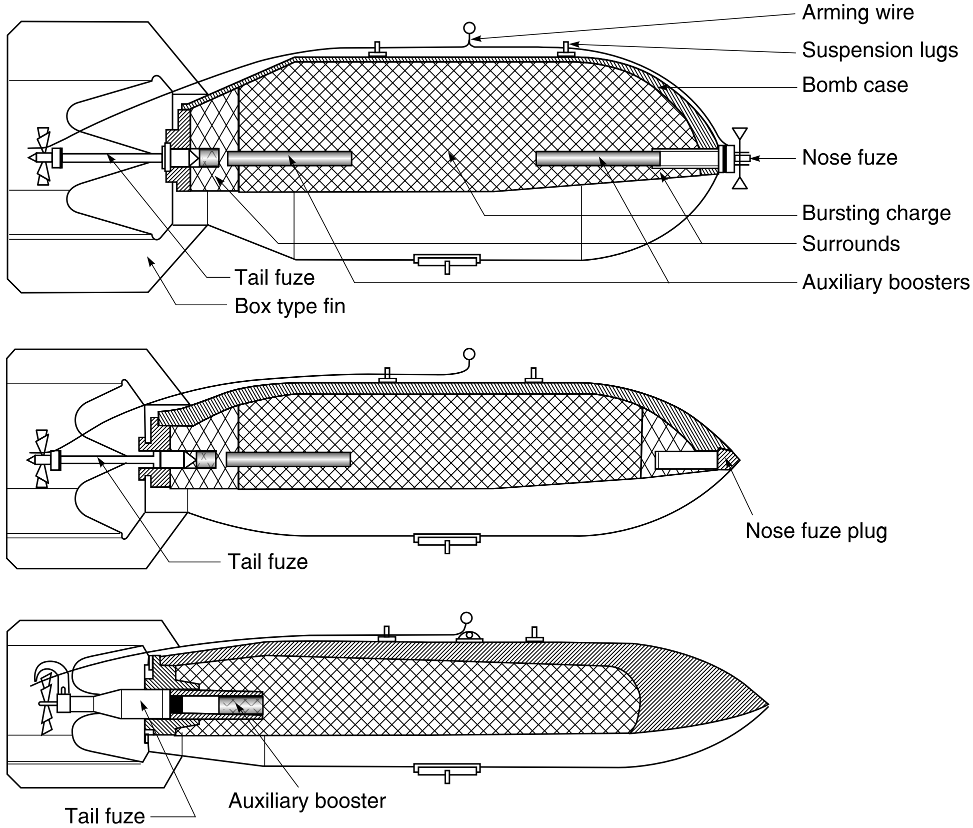


Figure 2. U.S. Army Air Force 1000 lb (450 kg) bombs from World War II also are typical, although later bombs became more streamlined to better suit jet speeds. From top to bottom they are general purpose (British medium charge), semi-armor piercing (British general purpose), and armor-piercing types, with ratios of charge to weight of ~57 percent, ~32 percent, and ~14 percent, respectively. The armor-piercing bomb is 185 cm long. Larger light case (British heavy charge) bombs with ~75 percent explosive weight were also employed for maximum blast effect.

employed in World War I were shells carrying chemical agents such as mustard gas.

Experience in World War I revealed a number of defects that impaired projectile performance, prompting a wide variety of refinements of detail. One innovation was a more reliable type of time fuse, based on mechanical clockwork. World War I also brought about interest in aerial bombs. Initially these were adapted from shells, but specially designed models soon were introduced. Owing to the much lower stresses, bombs could have lighter cases and explosive charges that could exceed 50 percent of weight. Larger charge weight fractions decreased fragmentation, effective against personnel and matériel, but increased blast effect against structures. Bomb weights increased as airplane capacities grew. Incendiary bombs filled with magnesium or jellied petrol (gasoline) were widely employed against flammable targets in World War II.

World War II brought accelerated improvements to existing types of weapons as well as

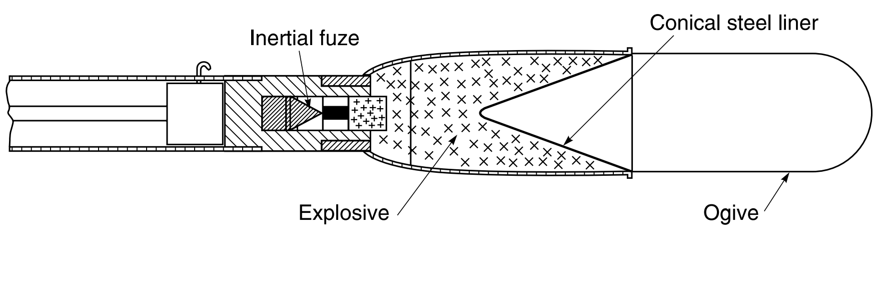


Figure 3. An early metal-lined shaped-charge munition, an American World War II Bazooka anti-tank rocket round of 2.75 in (7 cm) diameter.

innovation of new types. While TNT continued in very wide use, RDX (cyclotrimethylenetrinitramine) was introduced in quantity, frequently in mixtures with TNT. Often powdered aluminum was mixed with explosives. Both increased explosive power. An important development was introduction of metal-lined shaped charges. If a block of explosive has a conical cavity and if the detonation wave proceeds from the apex of the cone then the explosion products emerging from the cavity's walls are focused in a narrow stream along the axis of the cone. If the cavity is lined with metal, a narrow stream of metallic particles is projected. The tip of the stream reaches velocities of more than 5 kilometers per second (km/s), with bulk of metal mass following at more than 2 km/s. The pressure exerted by the jet is far beyond the yield strength of armor material, causing it to flow away like a fluid. In this manner the shaped charge digs a narrow hole in the armor approximately five to seven times as deep as its diameter, somewhat as a high-pressure jet of water in a block of gelatin.

Another wartime innovation was the proximity fuse, developed by the U.S. and Britain and not available to the Axis. This approximately trebled the effectiveness of antiaircraft fire and improved fragmentation and blast effects against ground targets. It worked using a tiny radar set carried in the nose of a shell or bomb. Cluster munitions emerged at this time as well. These were essentially grenades carried in bomb cases and scattered to increase the area covered by fragments.

The years following World War II brought the introduction of plastic-bonded explosives and further refinement of ammunition. Scientific research into the basic physics of weapons played an increasing role and led to development of a wide variety of munitions tailored to specific needs. Dueling tanks increasingly employed long-rod penetrators taking the form of slender darts fired to velocities approaching (by century's end) 2 km/s using light sabots to carry them down the bore of the large tank cannon. Made from a dense metal such as depleted uranium or tungsten, a rod can bore through steel armor of thickness approaching its length—approximately 60 cm for a 120-mm tank cannon, or nearly twice what could be achieved with conventional armor-piercing ammunition.

In the 1970s, explosively formed projectiles (EFPs) were developed for missile warheads. These consisted of shaped charges with shallow cones that ejected a compact mass of metal at velocities up to around 2.5 km/s. An EFP warhead detonated from a range of around 100 meters can deliver an effective armor-piercing blow. Another development of this period was the fuel–air explosive weapon (FAE), operating on the same principle as the detonation of fuel in a piston engine cylinder or of dust in a grain storage facility. Early FAEs employed a hydrocarbon-based fuel, which was dispersed in a cloud by one explosive charge and then ignited 100 milliseconds later by a second. The resulting deflagration (very rapid burning) typically produced a pressure wave less intense but lasting much longer than that of a high-explosive detonation. This long pressure pulse combined with the high energy content of the fuel (which, unlike an explosive, was able to take its oxygen from the air) to produce severe damage to structures over a wide area, as well as rupturing human organs. Effects were quite sensitive to environmental conditions, however, and efforts to develop FAEs with more predictability and flexibility were underway at century's end. Themobaric and enhanced blast weapons operated on similar principles; distinctions of terminology were not clearly settled. A variety of higher-energy explosives were investigated but considerations of cost as well as safety delayed their introduction.

See also **Battleships; Explosives, Commercial; Tanks; Warfare, Mines and Antipersonnel Devices; Warplanes, Bombers**

WILLIAM D. O'NEIL

Further Reading

Batchelor, J. and Hogg, I. *Artillery*. Charles Scribner's Sons, New York, 1972. General history of modern artillery.

Carleone, J, Ed. *Tactical Missile Warheads*. AIAA, Washington, 1993. Has introductory chapter on weapons history.

Frieden, D.R., Ed. *Principles of Naval Weapons Systems*. Naval Institute Press, Annapolis, 1985.

Padfield, P. *Guns at Sea*. St. Martin's Press, New York, 1974.

Wall, R. The devastating impact of sensor-fuzed weapons. *Air Force*, 81, March 1998.

White, M.P., Ed. *Effects of Impact and Explosion*. National Defense Research Committee, Washington, 1946. Particularly comprehensive on World War II bombs.

Warplanes, Bombers

Bombers apply aerospace technology to defeat an enemy through destruction of his will or ability to continue the conflict. In the twentieth century, the U.S. and the U.K found bombing particularly attractive because they were leaders in aerospace technology and disliked mobilizing large armies

and suffering heavy casualties. Bombing requires aircraft that can carry sufficient bomb loads over great distances, penetrate enemy defenses, find targets in darkness and poor weather, and bomb accurately. Effective campaigns require adequate bases, trained personnel, fuel, munitions, replacement aircraft, spare parts, and the intelligence capability to select and assess damage to the proper targets.

Aircraft technology received a tremendous boost during World War I. Until 1914, aircraft were essentially hand-built, but in 1918 France, Britain, and Germany mass-produced sturdy airframes and powerful, reliable, lightweight engines. Strategic bombing emerged as a theoretical way to break the stalemate of trench warfare, but existing aircraft could not strike military or industrial targets effectively. Zeppelin and Gotha raids on Britain did not break British morale, but profoundly influenced postwar strategic thought.

From 1918–1939, rapidly advancing technology dramatically increased speeds, ranges, and payloads. Aircraft evolved from fabric-covered biplanes to aerodynamically efficient, metal-skinned monoplanes with enclosed cockpits, dependable instruments, and retractable landing gear. Interwar theorists asserted that airpower would win future wars, but failed to appreciate the difficulties of navigation, target identification, bomber self-defense, and the size and number of bombs needed to cripple an enemy economy. Thus, no air force was truly ready for war in 1939, and the Anglo–American bombing offensive only fully matured in 1944.

The 1930s Royal Air Force (RAF) failed to create technologies to enable bombers to survive in daylight or to find, hit, and destroy precise targets at night. The RAF had little alternative to night area attacks on cities until late 1943, but persisted in this approach long after night precision bombing became possible. Germany created formidable defenses against night bombers, and after smashing these in mid-1944, Britain achieved the prewar dream of relentless, devastating air attack. By then, the RAF had perfected navigation and bombing aids (*Gee*, *Oboe*, and *H2S*), radar countermeasures (chaff and jamming), and an elite Pathfinder Force to mark targets for the main bomber force. The late-war strategic force consisted largely of Lancasters, whose four Merlin engines delivered a 6000-kilogram payload 12,670 kilometers, and Mosquitoes, an innovative wooden design that used speed, maneuverability, and altitude to escape the enemy. Unarmed Mosquitoes precisely delivered 1800 kilogram bombs over 2300 kilometers. Mosquitoes also performed ground support, torpedo bombing, night fighter, photoreconnaissance, and electronic warfare missions.

The Boeing B-17 equipped with Norden bombsights could, in 1934, fly faster than any available fighter, and American theorists concluded that such bombers could conduct unescorted daylight precision attacks. In 1943, however, unescorted B-17s suffered severe losses, forcing America to develop long-range fighters. B-17s and B-24s attacked oil, transportation, and industrial targets; but smoke, bad weather, and enemy defenses frustrated the effort significantly. Nevertheless, bombing tied down enormous defensive forces and crippled Germany's economy in late 1944. B-29s—developed during the war—incorporated then-revolutionary designs for wing trusses, pressurized cabins, and remote-controlled guns. B-29s based in the Marianas incinerated Japanese cities from low altitude at night and mined Japanese waters. This ruined Japan's economy before atomic attacks ended the war.

In the 1930s, Germany developed a strategic bombing doctrine and a twin-engine bomber force equipped with excellent navigational and blind-bombing aids, but Hitler misused this force against both Britain and Russia. Before the war, lack of sufficiently powerful, efficient engines hindered heavy bomber development, but more importantly, Germany failed to develop an effective long-range escort fighter. Wartime research led to such major advances as swept wings (which improve handling, stability, and control), turbojets, rocket engines, ramjets, turboprops, and prototype intercontinental bombers like the Me-264 and Ju-290. However, Germany did not produce advanced designs like the Me-262 and Ar-234 in adequate numbers soon enough to affect the war's outcome. After the war, Americans and Soviet aircraft greatly benefited from German airframe, propulsion, and aerodynamic research.

During the Cold War, the Americans, British, and Soviets developed heavy bombers, because initially only bombers could deliver nuclear weapons to enemy territory. Bombers lost their primacy with the development of intercontinental-range ballistic missiles (ICBMs) and air defense missiles in the late 1950s, but remained a flexible, highly accurate force that—unlike ICBMs—could be recalled or retargeted after launch. Jet engines and improved aerodynamics greatly enhanced performance, and initially led to development of high altitude, high-speed bombers. Increasingly effective networks of radars, interceptors, and surface-to-air missiles (SAMs) soon forced bom-

bers to operate at low altitudes and relatively low speeds. Bombers carried air-launched cruise missiles (ALCMs) to penetrate enemy defenses, and used electronic warfare to jam or deceive enemy sensors. Later bombers used "stealth" to reduce radar and infrared signatures. Each new generation of bombers was more survivable and lethal, but more expensive and purchased in smaller numbers.

Developed during World War II, Convair's B-36 never saw combat. After the war, the B-36 was the only American bomber with sufficient range and payload to strike Russia from North America with atomic weapons. Some favored canceling the B-36 and waiting for jet bombers, but the B-36 survived. With six propeller engines, B-36s flew too slowly to escape enemy fighters, and designers tested several concepts (such as "parasite" fighters carried in bomb bays) to reduce this vulnerability. The ultimate solution was a new bomber, but 385 B-36s served from 1948 to 1959.

The first American pure jet bomber, Boeing's B-47, was a revolutionary design. B-47s had six pylon-mounted jet engines, thin swept wings, and bicycle-type landing gear mounted in the fuselage. Some B-47s were equipped with rocket pods to shorten take-off length. B-47s were the first U.S. bombers that routinely used in-flight refueling, and also deployed at forward bases near the USSR. 2041 B-47s were built, serving from 1952 to 1966 in bomber and reconnaissance versions.

Boeing B-52s resembled larger, heavier B-47s, with swept wings, eight engines on pylons, and bicycle landing gear that pivoted for crosswind take-offs/landings. 744 B-52s were produced from 1954–1962 and subsequently modified extensively. Soviet SAMs required B-52s, originally designed for high altitudes, to fly low and carry up to twenty nuclear ALCMs. B-52s have dropped conventional unguided and precision munitions on Southeast Asia, Iraq, Yugoslavia, and Afghanistan. Eighty-five B-52s remain in the active force.

From 1946 until the late 1950s, the U.S. sought a delta-winged, high-altitude supersonic bomber. Supersonic flight created difficult problems of aerodynamic heating, structural fatigue, and flight control, which Convair's B-58 solved with composite materials and an innovative fuselage shape. The B-58 broke numerous speed records and won many aviation awards, but had unreliable bombing and navigation systems. More importantly, the B-58 was too vulnerable to Soviet SAMs, and was structurally unsuitable for low-altitude operations. Only 116 B-58s operated from 1961 to 1970. North American's XB-70 was a delta-winged Mach 3 nuclear bomber designed in the late 1950s that would have been the fastest and highest-flying bomber ever built. Like the B-58, the XB-70 would have been vulnerable to Soviet SAMs and unsuited to low-level operations. The XB-70 was cancelled after two prototypes were built.

Rockwell's B-1 emerged from the Advanced Manned Strategic Aircraft studies of the early 1960s. The B-1's variable geometry ("swing") wings enabled good high- and low-altitude performance and short take-offs/landings. Four turbofan engines permitted high subsonic and supersonic dash speeds. The B-1 had reduced radar signature and automatic high-speed terrain-following capability. Originally a nuclear bomber armed with bombs and up to 24 ALCMs, the B-1B was converted to a conventional precision bomber and fought in Iraq (in 1998), Yugoslavia, and Afghanistan. The B-1 entered service in 1985, and 93 remain in the inventory.

Northrop Grumman's B-2 "stealth" bomber employs special shapes, materials, and coatings to reduce visual, acoustic, infrared, and radar signatures. The B-2 resembles a "flying wing," and requires computer control for aerodynamic stability. America originally planned to acquire 132 B-2s for nuclear missions against the USSR, but scaled back to 21 aircraft when the Cold War ended. B-2s based in Missouri dropped precision conventional munitions in Yugoslavia and Afghanistan, and can reach anywhere on earth with one refueling from Missouri, Diego Garcia, or Guam.

The U.S. Air Force plans to upgrade its existing bombers until 2037. Future bombers will need sufficient range to strike anywhere on earth without refueling from U.S. bases. They may have "active stealth" and "quiet supersonic" capabilities. They will deliver nuclear, precision conventional, or directed-energy weapons to destroy fixed and mobile targets. Farther ahead, completely unmanned systems and "trans-atmospheric vehicles" (TAVs) may replace traditional bombers. TAVs would launch into low earth orbit, strike terrestrial targets with directed energy or hypervelocity kinetic projectiles, and land on a conventional runway.

Postwar British bombers emerged from Air Ministry specifications of December 1946, which called for a jet bomber able to carry a 4500-kilogram payload 6000 kilometers at 900 kilometers per hour and at 15,000 meters. Vickers, Handley Page, and Avro submitted designs, each featuring four jet engines mounted in the wing root, and from 1947 to 1964, Britain built 136 Vulcans, 86 Victors, and 107 Valiants, including prototypes.

Like the B-47, the Vickers Valiant had swept wings set high to permit a large bomb bay. The Valiant had a better payload and ceiling than the B-47, but the B-47 was faster. The Valiant became operational in 1955, and saw conventional action against Egypt in 1956. In 1962, the RAF realized that high-altitude bombers could not penetrate Soviet air defenses, and ordered the Valiants to fly at low level. The Valiant fleet quickly deteriorated due to the increased stress of low-level flying, forcing the RAF to withdraw them from service in January 1965.

The Handley Page Victor featured a T-tail, crescent wings, and the use of unique materials (a metal sandwich with a honeycomb filling) to reduce weight. The Victor became operational in 1958. In the early 1960s, Victors converted to low-level flight to penetrate Soviet defenses, and carried the Blue Steel nuclear ALCM. Due to the unexpected retirement of the Valiants, the RAF converted its Victors into tankers in the late 1960s. Victor tankers served until 1993, fighting in the Falklands and the Gulf War.

The Avro Vulcan was the world's first delta-wing bomber, and had a "kinked" leading edge to reduce buffeting. The Vulcan's delta wing, strong internal structure, and lightweight skin permitted a large payload and excellent, "fighter-like" handling at all speeds. Vulcans entered service in 1957, and gained improved electronic countermeasures (ECM), more powerful engines, larger wingspan, and in-flight refueling capability in 1960. In the early 1960s, the Vulcan converted to low-level flight and carried the Blue Steel ALCM. Vulcans incorporated terrain-following radar after 1966. Vulcans were intended to carry the American Skybolt air-launched ICBM, but America cancelled Skybolt in 1962. After British Polaris submarines entered service in 1969, Vulcans became conventional strike platforms, flying runway cratering and anti-radar missions during the Falklands War. The Vulcans retired in 1986.

Army officers dominated the 1930s Soviet military establishment, and many senior Air Force officers, aircraft designers, and engineers perished in Stalin's purges. During World War II, the Soviet Air Force focused on supporting the Red Army, and therefore in 1945 the Soviets lagged far behind Britain and the U.S. in strategic bombing theory, experience, and technology. Despite these obstacles, the Soviets created a strategic bomber force from scratch.

The Soviets initially copied the B-29, but the resulting Tu-4 lacked the range and survivability to attack the U.S. The twin-turbojet Badger, which replaced the Tu-4, was the first Soviet swept-wing bomber. Around 1500 Badgers were produced from 1953–1963 and some served as tankers and theater bombers until 1993.

Underpowered, inefficient jet engines hindered Soviet bomber development in the 1950s. The Bison, with four turbojets, could not reach the U.S. unrefuelled or carry ALCMs, so only 93 were built. The Tu-95, the world's only swept-wing turboprop bomber, had excellent range, but was slow and vulnerable. Nevertheless, roughly 100 Tu-95s comprised the primary element of the Soviet strategic bomber force from 1957 onwards. After 1959, Tu-95Ks carried one Kh-20 ALCM, upgraded to one or two Kh-22 ALCMs in 1982. In 1981, the Soviets began producing the Tu-95MS, and 63 (carrying 6 or 16 Kh-55 ALCMs) remain in service today.

In 1959, Khrushchev declared that ICBMs would be the primary strategic strike system, and transferred significant resources from bomber to missile production. In the 1960s and 1970s, the Soviets focused on developing ALCMs to penetrate Western air defenses, and also produced two theater bombers. The Soviets intended the Blinder to be a supersonic replacement for the Badger, but only built 300 Blinders due to unreliability and poor engine performance. The Backfire's variable-geometry swept wings and twin turbofans enabled supersonic dash speeds and low-altitude subsonic cruising. Some 500 Backfires were built after 1971, and saw combat in Afghanistan and Chechnya, carrying 3 Kh-55 ALCMs, 10 Kh-15 ALCMs, or conventional bombs. Backfires proved contentious during 1970s arms control negotiations. The Soviets agreed to remove the Backfire's refueling gear after the Americans insisted that Backfires, which could reach the U.S. with refueling, should be considered strategic bombers.

In 1970, the Soviets began designing the Blackjack, a variable-geometry supersonic intercontinental bomber carrying 12 Kh-55 ALCMs or conventional bombs. The Blackjack closely resembled the American B-1, and like the B-1, had reduced radar and infrared signatures. Fifteen Blackjacks entered service after 1987, and will probably remain in service until 2020. Russia recently cancelled rumored stealth bomber development for lack of funds, but continues to develop and deploy new precision ALCMs like the Kh-101 and Kh-555.

During the Cold War, the U.S. and Britain maintained strong navies to control the oceans, but the Soviets only needed the capability to deny NATO this control. Thus, maritime attack was a

primary mission for Soviet bombers, unlike for British and American bombers. Soviet bombers carried long-range, supersonic antiship cruise missiles (ASCMs) to attack NATO shipping outside the range of carrier air defenses. Satellites, reconnaissance Tu-95s, and other sensors provided targeting information for ASCM-equipped Badgers, Blinders, and Backfires from the 1950s until today.

Bombers never became the "cheap" war-winning weapons that theory proposed, because bombers needed constant improvement to penetrate defenses that were themselves constantly improving to counter bombers. Bombers played a crucial role in maintaining nuclear deterrence from 1945–1991. The advent of stealth and precision-guided munitions give bombers their current great utility, but defenses will doubtless eventually evolve to counter these technologies.

See also **Aircraft Design; Warfare, High Explosive Shells and Bombs; Warplanes, Fighters and Fighter Bombers**

JAMES D. PERRY

Further Reading

Brookes, A. *V-Force: The History of Britain's Airborne Deterrent*. Jane's, London, 1982.

Dorr, R.F. and Peacock, L. *B-52 Stratofortress*. Osprey Aviation, Oxford, 2000.

Futrell, R.F. *Ideas, Concepts, Doctrine: Basic Thinking in the United States Air Force, 1907–1984,* 2 vols. U.S. Government Printing Office, Washington D.C., 1989.

Jenkins, D. *B-36 Peacemaker*. Airlife Publishing, Shrewsbury, 1999.

Knaack, M. *Post-World War II Bombers, 1945–1973*. U.S. Government Printing Office, Washington D.C., 1988.

Levine, A.J. *The Strategic Bombing of Germany, 1940–45*. Praeger, Westport, CT, 1992.

Maurer, M. *Aviation in the US Army, 1919–1939*. U.S. Government Printing Office, Washington D.C., 1987.

McFarland, S.L. *America's Pursuit of Precision Bombing, 1910–1945*. Smithsonian Institution Press, Washington D.C., 1995.

Miller, J. *Convair B-58 Hustler*. Aerofax, Hersham, 1985.

Pace, S. *B-1 Lancer*. Airlife Publishing, Shrewsbury, 1998.

Podvig, P. *Russian Strategic Nuclear Forces*. MIT Press, Cambridge, MA, 2001.

Warplanes, Fighters and Fighter Bombers

Although new as weapons, fighters played an important role in World War I. Early in the war, reconnaissance planes and bombers were joined by fighters whose task it was to engage the enemy in aerial combat. Light machine guns were synchronized to fire through aircraft propellers. It was the German firm of Fokker which developed the first effective synchronizing device; this gave the Fokker planes, agile monoplanes, superiority over the Allies comparatively slow and less maneuverable biplanes. Aircraft development was then marked by a continuous catching-up process between German fighters on the one hand and French and British fighters on the other. By mid-1916 the British DH-2 and FE-2b had outclassed the Fokkers in speed and rate of climb whilst the French Nieuport-11 and -17 and the Spad S. VII distinguished themselves by superior agility. Later in the year Germany managed to close the gap with biplanes of the Albatros-A series, but Britain rapidly caught up with the Sopwith F-1 "Camel" and France with the Spad S. XIII. Although the new Fokker D-VIII and D-VIII proved superior to the Camels and Spads they came too late and were too few. At that time another airplane with high potential did not play any military role: the Junkers D-1, built in 1918 and derived from the J-1 model of 1915, was an all-metal, duraluminum monoplane. It was fast and maneuverable but needed more time for development.

Compared with the biplanes of World War I, the all-metal piston engine monoplane fighters of World War II reached speeds of about 470 mph (750 km/h), thereby doubling the speed performance of First World War aircraft. The British Hawker Hurricane Mk-1, which left the assembly lines in October 1937, showed great versatility and was used as a night fighter, fighter–bomber and ground attack aircraft. Owing to its decisive role in the Battle of Britain the Supermarine Spitfire became even better known. More than most other airplanes of the time, the Spitfire (of which 40 different models were built) benefited from the application of recent aerodynamic research. Its greatest adversary was the German Messerschmitt Bf-109, which during its first years of service, can be considered the best fighter worldwide. It was complemented by the Focke Wulf-190 which went into service in July 1941, surpassed the Me-109 in several respects and, regarding speed and maneuverability, proved superior even to the Spitfire Mk-V. However, the best fighter of the period 1943–1944 was the North American P-51 Mustang. Equipped with British Rolls-Royce Merlin engines it was extremely advanced in structural aerodynamics. The U.S. Republic P-47 Thunderbolt, a heavy single-engine, single-seat fighter, was of equal stature with high speed and excellent rate of climb, combined with heavy firepower and great sturdiness.

During World War II the piston engines as used in fighters had reached their limits of technical

capability. Therefore in Germany and Britain jet engines were developed which surpassed piston engines in performance. In Germany the experimental Heinkel He-178 flew in 1939; in Britain the experimental model of a Gloster, the E-28/39, followed in May 1941. Air engine designers had, among other difficulties, to cope with the lack of high performance materials for turbine engines. In the autumn of 1944 the German jet fighter Me-262 became operational and, in spite of various problems, proved to be an exceptional, fast aircraft. But its operational use came too late to have any significant impact on the outcome of the war.

Research during World War II, especially in Germany, had shown that swept-back wings eased shockwave problems at high speeds. Important U.S. and Soviet aircraft developed shortly after the war, such as the Lockheed Sabre and MiG-15, had swept-back wings, and others adopted delta-wing layouts. Research and development in aerodynamics, structural engineering, materials science, and related fields led to the development of fighters and fighter–bombers with improved performance characteristics. In the 1950s and 1960s there was an emphasis on speed.

The best supersonic jet fighters from the late 1950s onward could generally fly at twice the speed of sound, had fast rates of climb, great maneuverability and heavy firepower. The prototype of the French Dassault–Breguet Mirage III flew in November 1956 and was capable of high- or low-level interception in all weathers, tactical reconnaissance, and could also carry nuclear armament. Apart from this it exhibited experimental vertical take-off and landing (VTOL) ability. The U.S. company McDonnell started the McDonnell F-4 Phantom II project in 1953 as a response to the U.S. Navy's request for a supersonic twin jet all-weather assault fighter. The F-4 went into service in 1961, and during the 1960s and 1970s, was considered the best fighter in the world.

In the 1960s several European nations and aircraft producers joined in aircraft development programs because cost proved too high for a single nation to proceed by itself. In 1965 British and French manufacturers set up the Sepecat Consortium to build the Jaguar, a tactical support aircraft; four years later, in 1969; British, Federal German, and Italian firms established the Panavia Consortium to produce the multirole combat aircraft (MRCA) Tornado. One of the advantages of a multirole aircraft is economical: an MRCA can perform different operational tasks ranging from ground attack to reconnaissance and interception. For most of the flight, the Tornado can be flown automatically by its avionics. One of its most important features is the adoption of the Turbo Union RB 199 three-spool turbofan, in which part of the thrust is obtained from a large diameter fan, which makes for high performance and improved economy.

Similar to the Tornado, the U.S. General Dynamics F-16 (Fighting Falcon) is a multirole combat fighter and attack aircraft. Chosen by the U.S. Air Force in 1975 it employs the concept of relaxed static stability, which makes the fighter extremely maneuverable. The F-16 is equipped with high performance flight controls signaled by electronic fly-by-wire systems; it can be regarded as probably the best, but definitely the most cost-effective, combat aircraft of the 1980s.

Besides large military aircraft producers like the U.S., the Soviet Union or some supranational European consortia, Sweden is an interesting example of a relatively small, neutral state which, on the basis of excellent technical know-how and production facilities, developed advanced military aircraft of its own. The best known are the Saab-Scania J-35 Draken, an all-weather interceptor with maximum speed exceeding Mach 2; the Saab-37 Viggen; and recently, the JAS-39 Gripen. With the Viggen the designers employed a canard configuration (a canard is a fixed or moveable foreplane located ahead of the main wing thus making the aircraft virtually a supersonic biplane), which renders the plane highly maneuverable and enables it to operate out of short airstrips. The Gripen, while not superior in performance to the Draken or Viggen, is relatively small and comparatively easy to fly and maintain.

The "Eurofighter" is an aircraft designed to replace fighters and fighter–bombers in service in Europe and elsewhere. Although it is highly agile and has other first-class performance characteristics, it is also very expensive. But it is still cheaper than the US F-22 Rapier, a multimission air superiority combat aircraft developed jointly by Lockheed and Boeing. Its main features are advanced stealth technology and a thrust-vectoring system that yields phenomenal agility, although the Russian fighters SU-27 and SU-37 are in the same league. The French Dassault Rafale is equal to the Eurofighter and the F-22. The U.S. JSF (Joint Strike Fighter) currently being developed will also incorporate short take-off and landing ability.

See also **Aircraft Design; Warplanes, Bombers; Warplanes, Reconnaissance**

HANS-JOACHIM BRAUN

Further Reading

Angelucci, E. *The Rand McNally Encyclopedia of Military Aircraft 1914 to the Present*. Rand McNally, Chicago, 1981.

Batchelor, J. *Fighter: A History of Fighter Aircraft*. McDonald, London, 1973.

Bilstein, R.E. *Flight in America: From the Wrights to the Astronauts*, 3rd edn. Johns Hopkins University Press, Baltimore, 2001.

Bättig, T. On Target: *Moderne Kampfflugzeuge, deren Erkennungsmerkmale, Bewaffnung und Einsatzweise*. E.S. Mittler & Sohn, Hamburg, 1998.

Gunston, B. and Taylor, M.J. *Jane's Encyclopedia of Aviation*. Crescent, New York, 1993.

Gunston, B. *An Illustrated Guide to Modern Fighters and Attack Aircraft*. Prentice Hall, New York, 1987.

Munson, K. *The Blandford Book of Warplanes*. Blandford Press, Poole, Dorset, 1981.

Richardson, D. *Modern Warplanes: A Technical Survey of World's Most Significant Combat Aircraft in Service*. Salamander, London, 1982.

Stanley, W.L. *Measuring Technological Change in Jet Fighter Aircraft*. Rand Corporation, report 2249, Santa Monica, 1979.

Warplanes, Reconnaissance

The Montgolfier brothers' balloon flights in 1783 took the first step toward the development of aerial reconnaissance. Military application of the new technology came quickly with France's war with the Austrians in the last decade of the eighteenth century, though the small balloon corps established as part of the Napoleonic army lasted only a few years. Both the Union and the Confederacy used tethered balloons to spot for their artillery during the American Civil War and by the end of the nineteenth century, European armies were adding hydrogen balloons to their inventories. Design of the Parseval–Sigisfeld balloon, colloquially known as the "Drachen" (dragon), departed from the spherical type used earlier in the U.S. by the addition of a wind sock to the bottom of the envelope to point the main bag into the wind for greater stability. This innovation reduced the bobbing that had nauseated the crews of round balloons and prevented them from staying aloft for meaningful periods of time. Albert Caquot, director of the technical division of the French Aviation Militáire during World War I, further refined balloon design with the addition of side fins. Balloon use during World War I offered a technological advantage over heavier-than-air craft in that the balloon could remain aloft indefinitely and could be linked by telephone directly to the artillery batteries for which it was spotting.

Airplanes offered their own advantages. The Blériot XIs, BE-2s, and Albatros-B and -C types that made up the squadrons that mobilized at the outset of the Great War were able to roam over the battlefield, which allowed them to take intelligence photographs or to range targets beyond the sight of balloon observers. Airplanes had a downside, however. Their pilots and observers were unable to carry on a real-time conversation with anyone on the ground. Until development of the airborne radio receiver, which would have to wait for the next war, they were forced to fire flares, signal with lamps, fly to headquarters or the nearest artillery battery to land and deliver their message, or drop a note in a weighted bag. Eventually transmitters became available, but their use required deployment of an aerial consisting of a lead weight at the end of a long wire, which made the aircraft as long and unwieldy as the light planes of today which tow advertising banners over sports stadiums. If the mission came under attack while the antenna was out, the chance of outmaneuvering the enemy and surviving the combat was not good.

The science of photography also took some time to develop. Early wartime photographs offered either perpendicular or oblique views and were taken with handheld cameras by crewmembers leaning out of their cockpits. The results were shaky and the odds were against getting a series of exposures that overlapped into a complete view of the enemy area—one that could be made into a useful map. That changed with the invention of the serial camera. The serial camera could be mounted on the floor of the cockpit with its lens poking through a hole in the floor. The camera took a continuous series of photos. All that was necessary for the production of a good map was for the pilot to fly in a straight line at a constant altitude, and while that was easier said than done in the presence of enemy aircraft and antiaircraft guns—the determined reconnaissance crew sometimes being required to break off several times to fight and then return to finish the mission—it freed the crew to concentrate on defending the aircraft and represented a considerable advance over earlier practice.

After World War I and during World War II, technological effort was aimed at putting the camera at higher altitudes, theoretically out of the ability of the enemy to reach and destroy it, and to further increase its operational effectiveness by allowing it to operate in the dark. This led to development of electrical heating apparatus that prevented camera shutters from being adversely affected by the cold at high altitudes and to the slit

camera that adjusted the speed at which film was fed through the camera to the speed of the aircraft, an advance that improved the production of maps of enemy territory. Nighttime operations were aided by aerial flash equipment designed by Harold Edgerton of the Massachusetts Institute of Technology, which provided not only well-lit scenery, but also frozen imagery of the target, an asset vital to effective bomb sighting.

With the opening of the Cold War, the mindset that kept pushing reconnaissance to increasingly high altitudes and greater speeds took on a new importance as it not only kept the camera out of the enemy's physical reach, but out of his legal and political reach as well. The Royal Air Force's first jet bomber, the Canberra, counted on both speed and altitude to keep it away from enemy fighters. These advantages were to prove useful to reconnaissance as well and the Canberra still serves in the RAF inventory. The quest for speed and height led ultimately to the two best-known Cold War reconnaissance aircraft, the U-2 and the SR-71. Capable of cruising at 740 kilometers per hour (km/h), with a range of 3540 kilometers, and a ceiling of 17,000 meters. (21,000 meters and above in the later models), the U-2 represented the cutting edge in aerial intelligence gathering until it was superceded by the faster and higher flying SR-71. The Blackbird pushed the altitude envelope to over 26,000 meters and was able to maintain speeds of Mach 3.2.

In the wake of such performance, the next logical step was to put cameras out of the Earth's atmosphere and to take man out of the cockpit altogether. Today's Earth-orbiting satellites are able to photograph objects the size of a pack of cigarettes and maintain ongoing surveillance of areas of interest every hour-and-a-half as they continually circle the globe. The price of exclusive reliance on such high-tech intelligence may have revealed itself in the wake of the 11 September 2001 attacks on the World Trade Center and the Pentagon, however, as interest is beginning to focus on technology that combines the advantages of the satellite with systems more immediately controllable by and useful to the soldier on the battlefield. The Predator, an unmanned reconnaissance aircraft that sends imagery through receiving systems portable enough to be contained in a trailer but sufficiently sophisticated enough to be capable of being linked by satellite to military and intelligence offices all over the world is a twenty-first century system that combines high performance with relatively low operational costs. It is well attuned to the needs of a modern military that emphasizes the combination of high performance and low risk.

See also **Aircraft Design; Warplanes, Fighters and Fighter Bombers**

JAMES STRECKFUSS

Further Reading

Angelucci, E. *The Rand McNally Encyclopedia of Military Aircraft 1914 to the Present*. Rand McNally, Chicago, 1981.

Bilstein, R.E. *Flight in America: From the Wrights to the Astronauts*, 3rd edn. Johns Hopkins University Press, Baltimore, 2001.

Gunston, B. and Taylor, M.J. *Jane's Encyclopedia of Aviation*. Crescent, New York, 1993.

Munson, K. *The Blandford Book of Warplanes*. Blandford Press, Poole, Dorset, 1981.

Waste Processing, *see* **Nuclear Waste Processing and Storage**

Wind Power Generation

Wind is essentially the movement of substantial air masses from regions of high pressure to regions of low pressure induced by the differential heating of

Figure 4. At the USDA-ARS Conservation and Production Research Laboratory in Bushland, Texas, wind turbines generate power for submersible electric water pumps that are far more efficient than traditional windmills (background).
[*Photo by Scott Bauer, ARS/USDA*].

the Earth's surface. This simplistic view belies the complexity of atmospheric weather systems but serves to indicate the origin of climatic airflow.

The first attempts to harness wind power for electricity production date back to the 1930s. In Germany, Honnef planned a monstrous five turbine, 20 megawatt (MW), wind tower, several hundred meters high, a far cry from the sleek aerospace wind turbine generators (WTGs) of today. The design of a large scale WTG is limited to one of two formats realistically held to have good prospects. First, the horizontal axis type descended from those encountered by Don Quixote and common until recently in the flat lands of Europe; and second, the vertical axis machines of which the Darrieus rotor is perhaps the most common. Of the two, horizontal axis machines predominate, although the vertical axis type has many positive attributes, not the least of these being simplicity. The advantages and disadvantages of the two types are summarized in Table 1.

Horizontal Axis Wind Turbines

To gain some idea of the basic parameters that are taken into consideration in windmill design, we must resort to some basic fluid mechanics. A windmill of the horizontal axis variety is similar to a propeller, but it takes energy from the fluid instead of imparting energy to it. The flow pattern for the windmill is the opposite of that for the propeller: slipstream widening as it passes the disc.

Table 1 Relative advantages and disadvantages of horizontal axis and vertical axis wind turbine generators.

Type of generator	Advantages	Disadvantages
Horizontal axis	Rotor adjustments through pitch angle control relatively simple Optimal aerodynamic blade design	Complex Large structure
Vertical axis	Independence of wind direction Simple construction	Most designs have no blade control to regulate output Poor efficiency Problematical dynamic behavior

Since

$$\text{Kinetic energy} = \tfrac{1}{2}\ \text{mass (m)} \times \text{velocity (v)}^2$$

then

$$\text{Power available} = \tfrac{1}{2}\ \rho A v \times v^2$$

where ρ = air density, A = area swept by rotor and v = wind velocity.

In practice the actual power gained from the wind is far less than this. To account for this a nondimensional coefficient Cp expressed as

$$Cp = (\text{Force on rotor}) = (\tfrac{1}{2}\ Av^2)$$

is introduced and termed the power coefficient. Thus,

$$\text{Power delivered} = Cp \times \tfrac{1}{2}\ Av^3$$

The power coefficient is a measure of the efficiency of a wind turbine, and from theoretical considerations it is possible to show that it cannot be greater than 0.593. Actual efficiencies for the traditional windmill with a small number of sail-like blades are commonly of the order of about 5 percent, although much higher efficiencies of the order of 35 percent are achievable today.

Perhaps the single most important component of a horizontal axis wind turbine is the rotor. The most significant design parameters for the rotor are:

- Number of blades.
- Rotor blade construction.
- Hub design.

The number of blades is essentially a compromise between conversion efficiency and cost. In general practice, on large machines it has been found that because of diminishing returns the extra expense of incorporating more than two blades is not justified by the gain in output.

The rotor must be designed for a life of 20 to 30 years and must be able to withstand the complex oscillatory stress patterns created by asymmetrical wind loading, shadow effects, and the not inconsiderable bending loads due to the weight of the blades (sometimes as much as 50 meters long). Various construction techniques have been adopted for the rotor blades ranging from laminated wood and other composites through to steel; the only untried accepted technique is that employed by the aircraft industry of riveted aluminum, which is expensive.

The major components of the system for converting mechanical energy into electricity are step-

up gearbox, generator, current collectors, connecting shafts, couplings and locking brake. These components are usually housed in a nacelle (wood or metal enclosure), although some WTGs transfer the mechanical power to the base of the tower via a shaft or hydraulic pump/motor system, thus allowing the weighty generator to be ground based.

Apart from special cases, for relatively large WTGs only the normal three-phase generators of the synchronous or asynchronous type can be considered. Synchronization with the public grid is a problem because of the variability of the wind, so although synchronous generators are available relatively inexpensively and are in the main fitted to WTGs, they are by no means trouble free.

The use of asynchronous generator would circumvent the synchronization problem by allowing the generator to take its lead from the system. However, there is a limit to the number of asynchronous generators that may be connected to a particular grid system, and the efficiency of such generators is lower than for the synchronous variety.

Vertical Axis Wind Turbines

Vertical axis wind turbines (VAWTGs) have not been completely eclipsed by the development of large-scale horizontal axis machines. Much interest has been stirred by the arrival of the Musgrove variable geometry VAWTG, which combines the advantages of a vertical format with the ability to control the power output via the rotor blades.

The Darrieus rotor, first used in 1925, has for the greater part been limited to small- and medium-size test plants. The concept involved two, three, or four airfoils fixed at top and bottom to a vertical rotor shaft and bowed out at the center. Restraint of the rotor speed to design limits is achieved by means of brakes both in the form of spoilers at the center of the airfoils and disk brakes on the output shaft.

The Musgrove design, while retaining all the advantages of the vertical axis format, eliminates the complex blades of the Darrieus machine and uses conventional airfoils instead. The rotors were designed making extensive use of carbon fiber-reinforced plastic with titanium fittings and have an expected life of 40 years. Increasing wind speed causes the variable geometry rotors gradually to collapse vertically effectively reefing the blades and controlling the speed of rotation.

To put wind power in perspective, it must be considered that although the gross U.K. potential for electricity production from this source is as much as 20 percent of the total current consumption at a load factor varying between 10 and 45 percent. Thus wind power might be more usefully envisaged as an energy displacer for more expensive plants.

See also **Electrical Energy Generation and Supply, Large Scale; Energy and Power**

IAN BURDON

Further Reading

Gipe, P. *Wind Energy Comes of Age*. Wiley, New York, 1995.

Leggett, J. *The Carbon War: Global Warming at the End of the Oil Era*. Penguin, London, 1999.

MacNeill, J.R., Ed. *Something New Under the Sun: An Environmental History of the Twentieth Century*. Penguin, London, 2001.

Righter, R.W. *Wind Energy in America: A History*. University of Oklahoma Press, Norman, 1996.

Useful Website

http://telosnet.com/win/20th.html

World Wide Web

The World Wide Web (Web) is a "finite but unbounded" collection of media-rich digital resources that are connected through high-speed digital networks. It relies upon an Internet protocol suite that supports cross-platform transmission and makes available a wide variety of media types (i.e., multimedia). The cross-platform delivery environment represents an important departure from more traditional network communications protocols such as e-mail, telnet, and file transfer protocols (FTP) because it is content-centric. It is also to be distinguished from earlier document acquisition systems such as Gopher, which was designed in 1991, originally as a mainframe program but quickly implemented over networks, and wide area information systems (WAIS), also released in 1991. WAIS accommodated a narrower range of media formats and failed to include hyperlinks within their navigation protocols. Following the success of Gopher on the Internet, the Web quickly extended and enriched the metaphor of integrated browsing and navigation. This made it possible to navigate and peruse a wide variety of media types effortlessly on the Web, which in turn led to the Web's hegemony as an Internet protocol.

While earlier network protocols were special purpose in terms of both function and media formats, the Web is highly versatile. It became the

first convenient form of digital communication to have sufficient rendering and browsing utilities to allow any person or group with network access to share media-rich information with their peers. It also became the standard for hyperlinking cyberspace multimedia, or cybermedia, connecting concept to source in manifold directions and identifying them primarily by uniform resource locators (URLs), which became network-wide addresses.

In a formal sense, the Web is a client-server model for packet-switched networked computer systems defined by the protocol pair hypertext transfer protocol (HTTP) and hypertext markup language (HTML). HTTP is the primary transport protocol of the Web, while HTML defines the organization and structure of the Web documents to be exchanged via hyperlinks. HTTP and HTML are higher-order Internet protocols specifically created for the Web. In addition, the Web must also utilize the lower-level Internet protocols, Internet Protocol (IP) and Transmission Control Protocol (TCP). The basic Internet protocol suite is thus designated TCP/IP. IP determines how datagrams will be exchanged via packet-switched networks while TCP builds upon IP by adding control and reliability checking.

The Web can be thought of as an extension of the digital computer network technology that began in the 1960s. Localized, platform-dependent, low-performance networks became prevalent in the 1970s. These local area networks (LANs) were largely independent of, and incompatible with, each other. In a quest for technology that could integrate these LANs, the U.S. Department of Defense, through its Advanced Research Projects Agency (ARPA) funded research in internetworking (i.e., interconnecting LANs via a wide area network (WAN) which resulted in the first national network, ARPANET. For most of the 1970s and 1980s ARPANET served as the primary network backbone for interconnecting LANs for both the research community and the U.S. government. The open architecture and being built upon a robust, highly versatile and enormously popular protocol suite, TCP/IP, resulted in rapid growth and the gradual evolution into the Internet.

The Web was conceived by Tim Berners-Lee and his colleagues at CERN (now called European Laboratory for Particle Physics) in 1989 as a shared information space that could support collaborative work. Berners-Lee defined HTTP and HTML at that time and as a proof-of-concept prototype developed the first Web client navigator–browser in 1990 for the NeXTStep platform. Nicola Pellow developed the first cross-platform Web browser in 1991 while Berners-Lee and Bernd Pollerman developed the first server application—a phone book database. The Web began as a text-only interface, but the NCSA Mosaic browser added a graphic interface by the early 1990s.

Despite the original design goal of supporting collaborative work, Web use has become highly variegated. The Web has extended into a wide range of products and services offered by individuals and organizations—for commerce, education, entertainment, "edutainment," and even propaganda. Most Web resources are still set up for noninteractive multimedia downloads with the dominant Web theme being static HTML documents and noninteractive animations.

Web technologies have evolved beyond the original concept. The support of the common gateway interface (CGI) within HTTP in 1993 added interactive computing capability to the Web. An important use of CGI has been the processing of CGI forms, which enable input from the Web user-client to be passed to the server for processing. Forms were added to HTML around 1994 and allowed users to give feedback through standard graphic user interface (GUI) objects (e.g., text boxes, check boxes, buttons). Another technological advance began in 1994 with "helper apps:" extensions of the network browser metaphor, which diminished the browser-centricity by supporting multimedia through separate, special-purpose "players." In this way, a wider range of multimedia could be rendered than could be economically and practically built into the browser itself. Web browsers now include generic launchpads that could spawn prespecified multimedia players based on the file type/file extent. This generic, browser-independent approach would be challenged twice in 1996, first by "plug-ins" and then by "executable content." Plug-ins, as the name implies, are external applications which extend the browser's built-in capability for rendering multimedia files. Unlike helper apps which rendered the multimedia in an external window, plug-ins render the media within the browser's window in the case of video, or with simultaneous presentation in the case of audio. In this way the functionality of the plug-in is seamlessly integrated with the operation of the browser and as a result often proprietary and browser-specific because of this tight integration. While plug-ins proved to be a useful notion for creating extendable browsers, plug-in developers had to write and compile code for each target platform. This was eliminated through the notion of executable content, which maintained the tight integration between the

multimedia peruser and browser. The enabled browser will download the executable files that render the multimedia and execute them as well, all within the browser's own workspace on the client. This added a high level of animated media rendering and interactive content on the client side. There are several competing paradigms for Web-oriented executable content (e.g., scripting languages). However, executing foreign programs downloaded across the networks is not without security risk.

Several methods for dynamically creating web-page content have evolved; for example, dynamic HTML (DHTML) and server-side includes (SSI). Both server-push and client-pull technologies provide data downloads without user intervention. Server-push has been used to produce multiple-cell animations, slide shows, "ticker tapes," automatic pass-through of splash pages, and so on. A multitude of media-rich (if not content-rich) channels are available for use with this technology.

The World Wide Web represents the closest technology to the ideal of a completely distributed network environment for multiform communication. Perhaps the most significant impact of the Web will occur when it becomes a fully interactive, participatory, and immersive medium by default. Security and privacy issues will continue to emerge as new methods are integrated into current Web technologies. The secure socket layer is a security protocol that sits on top of TCP/IP to prevent eavesdropping, tampering, or message forgery over the Internet. Secure HTTP is a secure protocol over HTTP for identification when entering a server. However, it currently takes extra effort for Web users and servers to insure secure communication so needed by commonplace commerce. Although "cookies" were introduced to make the Web experience more useful by recording information about individual network transactions, they allow tracking of a user, seen by many as a loss of privacy.

The world will continue to become a smaller place as cultures continue to be only a click away. The Web promises to have one of the largest impacts on general society of any technology thus far created.

See also **Computer Networks; Electronic Communications; Internet; Packet Switching**

HAL BERGHEL AND YONINA COOPER

Further Reading

Berghel, H. Cyberprivacy for the new millennium. *IEEE Computer*, 34:1, 132–134, 2001.

Berghel, H. The World Wide Web, in *Encyclopedia of Computer Science*, 4th edn, Ralston, A., Reilly, E. and Hemmendinger, D., Eds. Nature Publishing Group, 2000, pp. 1867–1874.

Berghel, H. The client side of the Web, in *Encyclopedia of Library and Information Science*, Kent, A. and Hall, C.M., Eds. Marcel Dekker, New York, 1999.

Berghel, H. and Blank, D. The World Wide Web, in *Advance in Computers*, Zelkowitz, M.V. Academic Press, New York, 1999.

Berners-Lee, Tim. WWW: Past, present and future. *Computer*, 29,10, 1996, 69–77.

Comer, Douglas E. and Droms, R. E. *Computer Networks and Internets,* 2nd edn. Prentice Hall, Upper Saddle River, NJ, 1999.

Comer, D. *The Internet Book: Everything You Need To Know About Computer Networking And How The Internet Works*, 2nd edn., Prentice Hall, Upper Saddle River, NJ, 1997.

Comer, D. and Stevens, D. *Internetworking with TCP/IP,* 3rd edn. Prentice Hall, Upper Saddle River, NJ, 1996.

Hall, E. *Internet Core Protocols*. O'Reilly & Associates, Sebastopol, 2000.

Stevens, W. R. *et al.* . *TCP/IP Illustrated Volumes 1–3*. Addison Wesley, Boston, 1994.

Wright Flyers

By 1905 Wilbur and Orville Wright had developed, made, and flown the first practical airplane. This, their Flyer III, had a structure that withstood many take-offs and landings; it could bank, turn, fly figures of eight, and circle; and it was reliable enough to remain airborne for half an hour or more. The Wrights, despite having no formal training in engineering, had invested some six years of work in the production of this machine, and had combined observations of nature, experiments with wind tunnels, flying kites and gliders to construct their first Flyer. Each part of the Flyers—there were three in all—displayed the approach and originality of the Wrights, and their progression from early ideas to the eminently practical Flyer III can be marked out in four stages:

1. From 1896 to 1899 they studied the work of others and observed the flight of birds.
2. From 1899 to August 1901 they experimented with kites and gliders and validated their control systems, but found serious problems with existing knowledge.
3. From September 1901 to August 1902 they undertook intensive research on the form of airfoils, and reworking some aerodynamic problems allowed them to obtain a mastery over glider control.
4. From late 1902 to the winter of 1903 they concentrated on Flyer I, which carried an

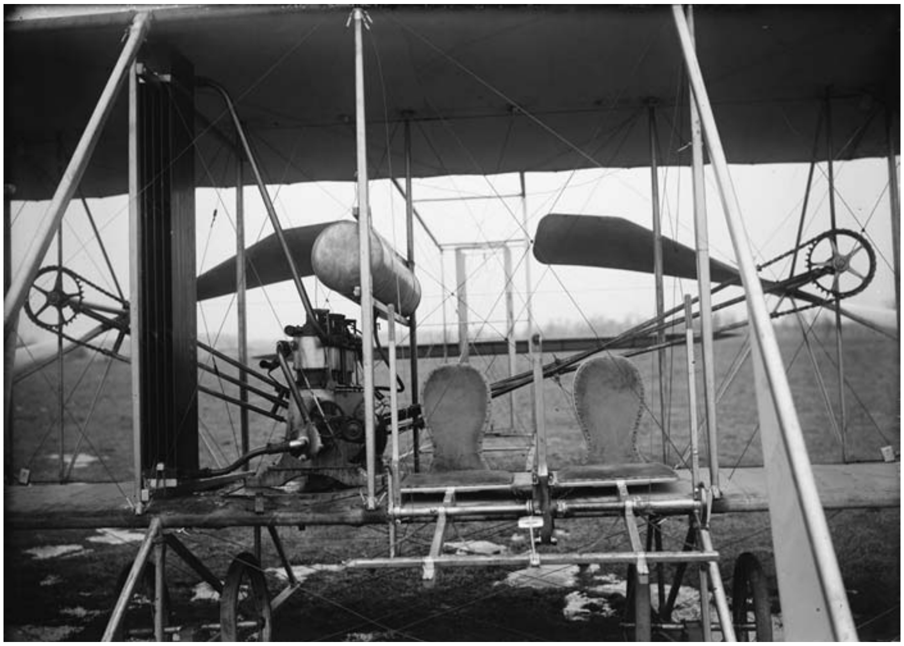


Figure 5. Close up of Wright brothers' airplane, including the pilot and passenger seats, 1911. [*Library of Congress, Prints & Photographs Division, LC-DIG-ppprs-00690*].

engine of their own design that drove two highly efficient propellers, again of their own design. It was this airplane that is credited as the first heavier-than-air machine to fly under its own power when Orville piloted it for 12 seconds on 17 December 1903.

The brothers quickly realized that a good system of control was essential in order to achieve flight, and they concluded that methods which relied on shifting the center of gravity of a machine were unsatisfactory. Any effective airplane must be controllable, or be stable, about the three axes of roll, pitch, and yaw; and it was the flight of birds that provided them with their ideas. Bird flight is in fact very complicated, but it is clear that the main control elements are the wings and the tail. Birds use combinations of movements in their flying, but the Wrights decided that the two most important elements were a twisting of the wings, and a raising and lowering of the tail. To test the effectiveness of wing twisting, or wing warping as they were to call it, a biplane kite equipped with a fixed elevator was built and flown in August 1899. The tests were successful, and wing warping was incorporated into a full-sized glider (No. 1) which they flew in October 1900—usually unmanned. The fixed tailplane was replaced with a front-mounted elevator. Initially the glider's wings were given upward slopes (dihedral), as this configuration confers a degree of stability in roll, but in fact the glider was difficult to control, and the Wrights made little use of dihedral in later designs. Their second glider (No. 2) incorporated lessons they had learned with No. 1. It was larger and well able to carry a pilot, and its wings were given downward slopes (anhedral). It flew reasonably well, but showed a tendency when banked to side slip, slew round and crash. The Wrights were led to doubt data on airfoils and control they had acquired from Otto Lilienthal in Germany and Octave Chanute in the U.S., and they decided to carry out and rely on their own investigations.

To find the best form of airfoils they used wind tunnels and horizontal force balances carried on the handlebars of a bicycle. Extensive series of tests led them to believe that they would be able to calculate the performance of an airplane in advance of its construction, for they could relate the geometry of the wing to airspeed and lift. The major aerodynamic problem to be solved was to find a means of counteracting the tendency of the planes to slew round, side slip, and crash in turns or when hit by gusts. A rear-mounted vertical fixed rudder was seen to be a solution, but this acted as a lever and aggravated the slewing tendency. Coupling the rudder with the wing-warping controls was a partial solution, and in fact it was found that as a result of overcompensating for slewing it was possible to make smoothly banked turns. Their glider No. 3 was the vehicle for these changes and enabled the Wrights to demonstrate the efficiency of their control system. They were almost ready to build a powered machine.

Two formidable problems had to be solved before a powered machine could be built—the provision of a suitable engine, and efficient propellers. To obtain both, the brothers relied on their own resources—they designed and built their own internal combustion, four-cylinder, 12-horsepower engine, and used their work on airfoils to design propellers. Flyer 1 was a biplane and incorporated all the lessons learnt during their experiments with gliders. Its propellers were contrarotating to counteract gyroscopic effects and were chain driven. The pilot lay flat on the lower wing, and operated an independently operated elevator, and a double rudder whose action was effected by the hip-cradle with which the pilot worked the wing-warping controls. Orville's 12-second, 37-meter flight took place at Kill Devils Hill, North Carolina, at 10:35 a.m. on December 17, 1903. Three more flights were made that morning; the last, made by Wilbur at noon, lasted for 59 seconds and covered 862 meters over the ground. These flights were the first in which a piloted machine had taken off under its own power, had flown under full control, and had landed on ground level with its take-off point. However, there were problems with control, and Flyer I was not a fully practical machine. Flyer II, similar to I, was used to learn more of the techniques of flying and to iron out some minor problems, but it tended to stall in tight turns. This difficulty was overcome in Flyer III, which first flew in June 1905. In this machine the Wrights had a practical airplane. Their solution was to decouple the rudder from the warp controls, enabling rotation about all three axes to be controlled independently, a situation that persists to this day. Like all their flying machines Flyer III was essentially unstable and had to be flown by the pilot, yet so good were their machines that they remained unchallenged by others until around the end of the Edwardian period in 1910.

See also **Aircraft Design; Internal Combustion Piston Engine**

COLIN HEMPSTEAD

Further Reading

Combs, H. *Kill Devil Hill: Discovering the Secret of the Wright Brothers*. Boston, 1979.

Gibbs-Smith, C. H. *The Wright Brothers: A Brief Account of their Work, 1899–1911*. London, 1963.

Hallion, R. *Wright Brothers, Heirs of Prometheus*. Washington, 1978.

Jakab, P.L. *Visions of a Flying Machine*. Shrewsbury, 1990.

Kirk, S. *First in Flight: The Wright Brothers in North Carolina*. Winston-Salem, NC, 1995.

Scott, P. *The Pioneers of Flight: A Documentary History*. Princeton, NJ, 1999.

Westcott, L. *Wind and Sand: The Story of the Wright Brothers at Kitty Hawk*. New York, 1983.

Unitt, P. Charlie Taylor and the Wrights' 1909 Military Flyer. *Am. Aviat. Hist. Soc.*, 43, 228–35, 1998.

X

X-Ray Crystallography

X-ray crystallography is a technique allowing for the determination of crystal structure. Throughout the twentieth century it has been used to determine increasingly complex structures, from the inorganic, through organic, to the biological and, most famously, the determination of the structure of DNA. It has been of tremendous utility in the materials sciences and in molecular biology. The knowledge of three-dimensional molecular structures can determine the unknown chemical formula of a compound, and it is critical for understanding biological processes such as how molecules interact, how enzymes catalyze reactions, and how drugs act. It is also a prerequisite for new drug design.

The technique relies on the fact that x-rays are waves with a wavelength of about 10^{-10} meters or 10 Å, which is comparable to intermolecular distances. X-rays can thus be diffracted by the periodic arrangement or lattice of molecules or atoms in a crystal. Following the discovery of x-rays by William Conrad Roentgen in 1895, the diffraction of x-rays by a crystal was discovered in 1912 by Max von Laue, for which he was awarded the Nobel Prize for physics in 1914. In 1913, the British physicist, William Lawrence Bragg, analyzed the diffraction pattern created on photographic film by simple crystals such as rock salt, reasoning what the three-dimensional crystal structure must be in order to create just that pattern. Using the x-ray spectrometer developed by his father, William Henry Bragg, he made a series of such determinations, from which he provided a general conceptualization of the relationship between crystal structure and the diffraction patterns created upon irradiation with x-rays of varying wavelengths. Bragg and his father received a Nobel Prize in 1915 for the analysis of crystal structure by means of x-rays. Since then, many improvements have been made both in the technique for the generation and detection of diffraction patterns and also in the reasoning from diffraction pattern to crystal structure.

During the first decade, progress was made in the basics of the technique, partly in the understanding of what intensities of x-rays led to what degree of blackening of the diffraction spots, and partly in the control of monochromatic and multichromatic x-rays. The development of x-ray tubes with higher intensities, especially of the hot-cathode variety, helped crystallography along. As in radiology, a problem continued to be posed by the heating of the x-ray tube's target, and many solutions for cooling were tried. In this early period it was also realized that one could analyze powders in addition to the single crystals initially used. Finally, the relationship between Fourier analysis and diffraction was being probed; the eventual conceptualization being that each single instance of diffraction corresponds to a Fourier component of the crystal's charge density. A Fourier transform is a mathematical operation that can be thought of as operating on the charge density: the observed diffraction pattern is the outcome. Taking the result and running it through an inverse Fourier transform gets back the original charge density.

In the 1920s this was taken up and developed into two-dimensional Fourier analysis. In the same decade, x-ray intensity was being standardized for the burgeoning discipline of radiotherapy, also yielding the possibility of absolute intensities for diffraction work. X-ray crystallography was now

mature enough to yield information about organic crystals—structural information agreeing with the structures determined by organic chemistry.

In the 1930s, much of the theory of x-ray crystallography was consolidated in concepts useful for the increasingly routine deductions. Fourier projections demand a large number of measurements.

In the 1930s, the technique of rotating the photograph replaced the use of the spectrometer because Fourier projections require a large number of reflection measurements. J. D. Bernal at the University of London provided the interpretation of rotation photographs. With this technique the structure of several organic crystals were determined, along with their bond length and bond angle. Linus Pauling at the California Institute of Technology prominently contributed to the general understanding of the nature of the chemical bond. Fourier analysis was imaginatively manipulated and became so pliable as to replace the original method: guessing structures and then working out the resultant theoretical diffraction pattern until it matched the one measured. Phase indeterminacy remained a general problem (both the amplitude *and* the phase of the diffracted waves are needed to compute the inverse Fourier transform), but techniques were worked out for individual cases. Important concepts of x-ray crystallography came into general use in this period, such as reciprocal lattices and Brillouin zones.

During World War II the development of x-ray crystallography was generally interrupted, but after 1945, it accelerated due to increased funding for science. The community of crystallographers was consolidated with the establishment of the International Union of Crystallography and the founding of the journal *Acta Crystallographica* in 1948. The determination of inorganic compounds was by now fairly routine and the race with inorganic chemists was on to determine ever more complex structures, for example strychnine. Crystallographers themselves found that they broke through a barrier in this period by surpassing organic chemists' knowledge with the determination of crystal structure and chemical formula of vitamin B12 in 1956 by Dorothy Hodgkin, who won the Nobel Prize for her crystallography work. Vitamin B12 was by far the largest molecular structure ever solved by x-ray crystallography at that time, and this discovery led the way to the synthetic development of B12.

In the second half of the twentieth century, x-ray crystallography has developed into the most prominent tool in structural biology. In the 1950s, Pauling and his colleagues determined the α-helix in protein structure, Watson and Crick (at the University of Cambridge, U.K.) used the measurements of Rosalind Franklin (working with Bernal in London) and others to determine the double-helix structure of DNA, and John Kendrew and Max Perutz (also at Cambridge) determined the structure of myoglobin and hemoglobin. The initial placing of heavy atoms within large molecules happened to solve the phase problem, and this was developed into a central technique. The experimental set-up was increasingly automated, yielding more and more data, and the so-called direct approach rendered the calculations amenable to electronic computers. Human conceptualizations remained important, but increasingly tasks in the inference from diffraction pattern to crystal structure were algorithmized.

While the demands of structural biology for x-ray crystallography have continued to grow dramatically for the last half century, the most important technical development has been the shift from x-ray tubes to synchrotron radiation (i.e., radiation that is produced by charged particles moving at relativistic speeds in a magnetic field). High-energy physics experiments required storage rings for electrons; and as these particles are accelerated centripetally through bending magnets, x-rays are produced.

Demand and supply have interacted closely; for example, in the early stages of the European Molecular Biology Laboratory (EMBL). Diffraction experiments at the Deutscher Elektron-Synchrotron (DESY), in Hamburg, Germany, on insect flight muscle in 1971 not only made an immediate impact on studies in structural biology but also provided a practical example of how an international molecular biology facility could foster novelty. John Kendrew, EMBL's first director, used this synchrotron research to bolster the demands for large-scale funding.

Initially, the interest focused on greater intensity of x-ray fluxes but as time went on, other properties of synchrotron x-rays (continuous spectrum and definite time structure) have also been utilized. New equipment and beam lines were added at many facilities during the late 1970s and early 1980s, providing increasing access to the radiation for small-angle scattering, high-resolution x-ray spectroscopy, and protein crystallography experiments. Computer specialists helped to improve technology and analysis. They worked to construct data acquisition systems and analysis packages for structural problems, most notably for fast time-resolved synchrotron radiation experiments. They

also developed early interactive computer graphics for x-ray crystallography and molecular modeling, a mini-revolution that replaced balsa wood and mechanical models. The demand for access to synchrotron radiation by visiting scientists intensified considerably. The demand led to the development of two-dimensional detectors and online imaging plate scanners for protein crystallography.

The experiments at DESY were parasitic, in the sense that they had to adapt a beamline for structural biology at a synchrotron that had been designed for high-energy physics experiments, such as particle collisions. X-rays produced at these physics facilities had originally been regarded as waste. In the 1980s, second-generation facilities were dedicated specifically to the production of synchrotron radiation. Area detector systems were developed for the new facilities, replacing photographic film. Several improvements relate to the increase in data processing power: for example, graphical display, such as the fitting of models onto electron density maps; and stereochemically restrained refinement, generally adopted through computer programs. The introduction of recombinant DNA technology was very significant for x-ray crystallography in this period because it enabled the production of samples of arbitrary molecules large enough to generate diffraction patterns.

In the second-generation facilities, it was discovered that magnetic array devices, termed wigglers and undulators, markedly enhanced the x-ray beams. The design of synchrotrons incorporating wigglers and undulators has led to a third generation of synchrotrons in the 1990s. Other important developments in the 1990s include charge-coupled device detectors, cryopreservation, multiwavelength anomalous diffraction (MAD) phasing, selenomethionyl proteins, and structure-solving automation. Since the turn of the century, MAD has become the method of choice.

See also **Cancer, Radiation Therapy; X-Rays in Diagnostic Medicine**

ARNE HESSENBRUCH

Further Reading

Bragg, W. L. *The History of X-Ray Analysis*. Longman, Greens & Company, London, 1943.

Ewald, P.P., Ed. *Fifty Years of X-Ray Diffraction*. Oosthoek's Uitgevermaatschappij, Utrecht, 1962.

Hendrickson, W.A. Synchrotron crystallography. *Trends Biochem. Sci.*, 25, 637–643, 2000.

Hessenbruch, A. A brief history of x-rays. *Endeavour*, 26, 137–141, 2002,

Michette, A. and Pfauntsch, S. *X-Rays: The First Hundred Years*. Wiley, New York, 1996.

Olby, R. *The Path to the Double Helix*. Macmillan, London, 1974.

X-Rays in Diagnostic Medicine

William Conrad Roentgen, a German physicist, discovered x-rays in November 1895. X-rays are a form of electromagnetic radiation, ranging in wavelength from approximately 10^{-9} to 10^{-12} meters, or 0.01 to 100 angstrom (Å). Unlike visible light, x-rays have sufficient energy to penetrate human tissues, and it is this property that makes them so useful in diagnostic imaging. Roentgen discovered x-rays while investigating discharge from an induction coil through a partially evacuated glass tube. The tube was covered in black paper and the room was in darkness, yet he noticed that a fluorescent screen in the room became illuminated. Other scientists had encountered similar phenomenon while experimenting with electrical discharges through gas filled tubes. Sir William Crookes, a leading scientist in the field, found photographic plates in his laboratory fogged but did not realize it was as a result of his experiments, and he returned them to the manufacturer. Roentgen, however, was the first to realize the significance of the effect, and he received the first Nobel Prize for physics in 1901.

X-rays, as opposed to visible light rays, penetrate matter because of their higher energy. This is a result of high-energy electrons generated in the x-ray tube; the energy of electrons incident on the target sets a maximum limit on the energy of the x-rays produced. The incident electrons gain energy from the high electrical potentials across the tubes. Up to the 1920s most x-ray tubes used induction coils (Figure 1). The inner coil is connected to the battery through a switch and an interrupter device. Once the switch is closed, current flows in the circuit, and the iron core becomes magnetized. The interrupter consists of a screw and sprung switch. The iron head of the sprung switch is attracted by the magnet and disconnects from the contact screw breaking the current path. The core becomes demagnetized, the sprung switch springs back to make contact with the screw, which once more allows current to pass, magnetizing the core. This process repeats at high speed causing a varying current in the inner coil, which induces alternating current of high potential in the secondary coil. This places a potential across the terminals T1 and T2, which are in parallel with the x-ray tube. The potential is sufficient for a spark to cross the gap between the two terminals. If the terminals are further separated, after a certain point the tube's

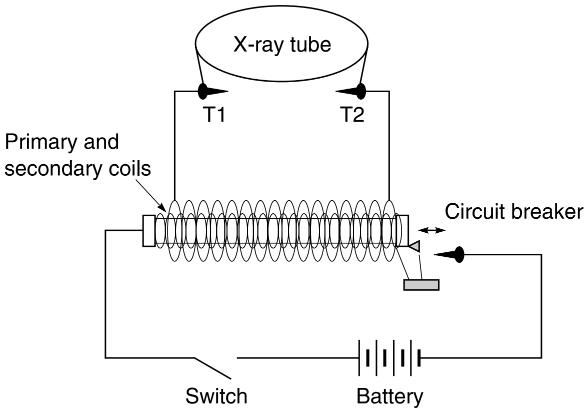


Figure 1. Induction coil.

resistance is less than the resistance between the terminals and the potential is available across the tube. The oldest method of specifying penetrating power of x-rays was by determining the length of the spark gap (the kilovoltage is equal to the equivalent spark gap in air).

The military were among the first to take advantage of x-ray technology to locate bullets and shrapnel in wounded soldiers. Power sources were a particular problem for x-ray systems in the field, however, and solutions ranged from gasoline engine-driven dynamos, to biplane engines in World War I, with even pedal power from a tandem bicycle used in the Sudan in 1898 (with the bike used for charging storage batteries). High-voltage transformers were introduced around 1919 and became the basis of high voltage generators in x-ray tubes.

In gas-filled tubes the potential across the tube causes ionization of the gas and electrons were attracted to the positive side of the tube. When the stream of electrons hit the target (originally the end of the glass tube), heat and x-rays were produced. Most of the energy was converted to heat, which placed a limit on the amount of time x-rays could be produced because the target would melt. The two most important early improvements in x-ray gas tube design were by Campbell Swinton, who introduced a sheet of platinum as a metal target, and by Herbert Jackson, who used concave cathodes to focus electrons onto a small area of the metal target producing more sharply defined images. The gas tube was unreliable, as x-ray production depended on the variable factor of the gas content. Richardson's discovery of thermionic emission in 1902 formed the basis for a major advance in x-ray tube design. W. D. Coolidge of the General Electric Company developed a thermionic x-ray tube (the Coolidge tube) in 1913 that used an almost perfect vacuum. Electrons are boiled off the cathode (thermionic emission) when current passed through the filament circuit. These electrons are accelerated across the tube producing x-rays when they strike the anode target. By 1915 Coolidge had developed rotating anodes, achieving target rotation of 750 revolutions per second, which effectively increased the area of the anode target and increased the anode's ability to dissipate heat. Modern x-ray tubes have most of the design features developed in the first 40 years of research; however, tubes generally have a selection of cathode filaments, anodes rotate at much higher speeds, the target is usually tungsten (which has twice the melting point of platinum), and improved designs for high-voltage generators and regulators have allowed a greater degree of control over x-ray energy (Figure 2).

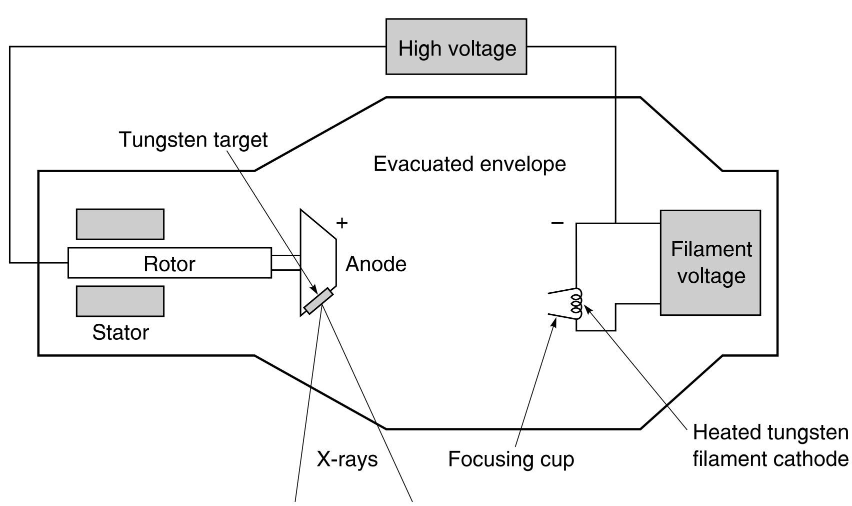


Figure 2. X-ray tube.

X-rays have a number of properties that allow their detection, and they affect a photographic emulsion in the same way as light. Films are placed behind the patient to capture the emerging x-rays (Figure 3). Screens, which reduce exposure requirements in both dose and time, have been used since the early days of x-rays. When x-rays interact with matter they may be scattered at different angles, and scatter disturbs the correspondence between points in the image and in the patient, resulting in reduced image contrast. Moving grids for scatter rejection were introduced in the 1920s. The grid blocked photons emerging from the patient at oblique angles to the detector. The grid was moved during exposures to prevent the grid lines from appearing on the image. Digital radiographic techniques, developed in the 1980s by the Fuji Corporation, introduced a photo-stimulatible phosphor screen as a primary image receptor. Use of digital detectors for radiography occurred in the 1980s and 1990s, although film remained the dominant detector type.

Another property exploited for the detection of x-rays is that they cause fluorescence in certain materials. By 1896, Thomas Edison had examined 1800 chemicals to detect and compare their x-ray fluorescent properties. His skiascope used a platino–barium cyanide fluoroscopic plate installed in a visor, which was held up to the eyes to allow observation of x-ray fluorescence. Fluoroscopic images are dynamic, with the illumination pattern on the screen responding to changes in the object being imaged. Fluorescence produced very low levels of illumination, and in many applications the eye had to be dark-adapted before images could be properly viewed. Image intensifiers, which were developed in the 1960s, greatly improved fluoroscopic imaging. These intensifiers convert the fluorescent light into electrons and accelerate them across to an output plate, which amplifies the signal. The output plate, being smaller, produces an additional geometric gain. The output screen is linked to a camera that allows display of dynamic images on a monitor.

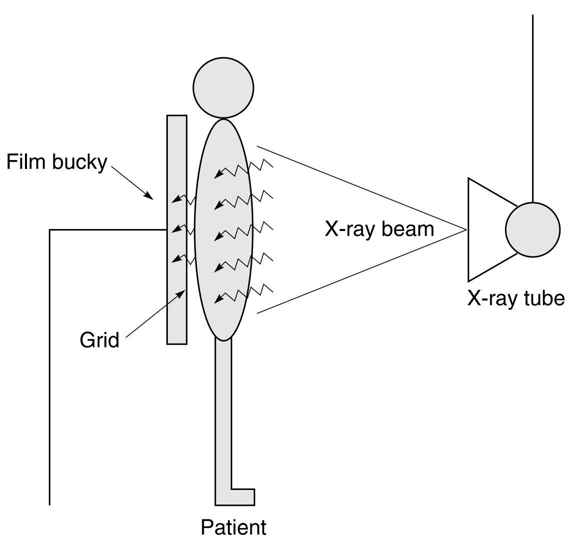


Figure 3. Basic chest x-ray radiography system.

Digital technology has revolutionized x-ray imaging. Digital subtraction angiography (DSA), which exploits the use of a contrast agent and digital imaging processing for use in vascular imaging, was developed in the early 1980s. The contrast, a dye with high x-ray attenuation properties, is injected into the blood stream. In DSA, images of a region before and after the dye is injected can be digitally subtracted. The resultant image clearly shows the vein (or area where contrast was injected), while the overlying bone structures, which might inhibit visualization, are subtracted.

In the first few years after the discovery of x-rays, it became apparent that there were hazards associated with their use. Accounts of the early years contain numerous reports of injuries and even the death of radiation workers. In 1904 Beck described three levels of Roentgen ray burns varying from itching symptoms to painful blisters and skin discoloration and ulcers. In 1915 the Roentgen society in London adopted radiation protection recommendations, and the first dose limits were proposed in 1925. The International Committee on X-ray and Radium Protection (ICXRP, later the ICRP) was formed in 1928. In 1954 the U.S. National Committee on Radiation Protection (NCRP) put forward the concept of ALARA, an acronym for "as low as reasonably achievable." Because x-rays are a form of ionizing radiation, all examinations must be clinically justified and doses kept "as low as reasonably achievable." The ALARA concept was stated in the first ICRP publication in 1959. Radiation protection recommendations continued to be reviewed and updated throughout the twentieth century in an effort to optimize safety in the use of diagnostic x-rays.

X-ray technology applications routinely used in hospitals include film screen systems, fluoroscopy systems that can provide dynamic image information, and computed tomography (CT), which provides high-contrast image slices through the patient. The many clinical applications include examination of broken bones, angiography pro-

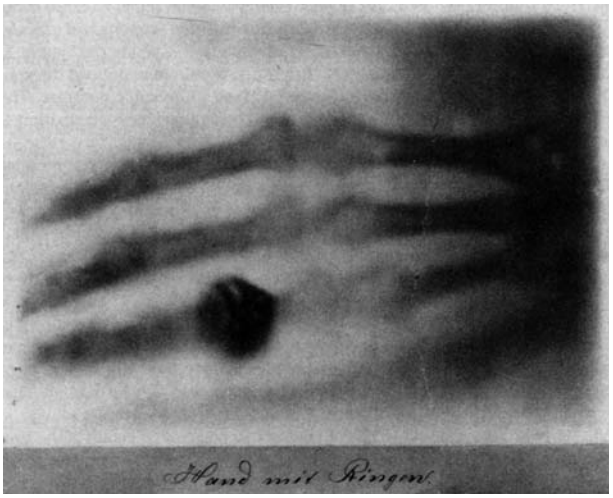


Figure 4. Hand of Mrs. Wilhelm Roentgen: the first x-ray image, 1895.

cedures, and identification of tumors, such as in mammography. In interventional procedures, x-ray technology is used to assist radiologists and surgeons in removing blockages in blood vessels and inserting pacemakers. High-energy x-ray systems may also be used therapeutically to provide lethal radiation doses to cancerous tumors. Future developments are likely to concentrate on digital systems, and many hospitals are implementing wholly digital departments. Picture Archive Communications Systems (PACS) allow images to be stored centrally and transmitted to display devices around the hospital. Instead of looking at x-ray films on light boxes, clinicians examine films displayed on high-quality monitors.

See also **Angiography; Cancer; Radiation Therapy; Particle Accelerators, Linear; Tomography in Medicine**

COLIN WALSH

Further Reading

Boone, J.M. X-ray production, interaction, and detection in diagnostic imaging, in *Handbook of Medical Imaging, Volume 1: Physics and Psychophysics*. Beutel, J., Kundel, H.L. and Van Metter, R.L., Eds. SPIE Press, U.S., 2000.

Brown, B.H., Smallwood, R.H., Barber, D.C., Lawford, P.V. and Hose D.R. *Medical Physics and Biomedical Engineering*. Institute of Physics Publishing, U.K., 1999.

Bushberg, J., Seibert, J.A., Leidholdt, E.M. and Boone, J.M. *The Essential Physics of Medical Imaging*, 2nd edn. Lippincott, Williams & Wilkins, Philadelphia, 2002.

Curry, T.S., Dowdey, J.E. and Murry, R.C. *Christensen's Physics of Diagnostic Radiology*, 4th edn. Lea & Febiger, Philadelphia, 1990.

Dance, D.R. Diagnostic radiology with x-rays, in *The Physics of Medical Imaging,* Webb, S., Ed. Institute of Physics Publishing, U.K., 1988.

Davidovits, P. *Physics in Biology and Medicine*, 2nd edn. Harcourt/Academic Press, U.S., 2001.

Mould, R.F.*A Century of X-Rays and Radioactivity in Medicine*. Institute of Physics Publishing, U.K., 1993.

Smith, F.A. *A Primer in Applied Radiation Physics*. World Scientific Publishing, Singapore, 2000.

Index

B

D

F

G

H

N

Q

R

S

T

U

V

W

X

Y

Z